TOWARDS UNDERSTANDING HOW MEMBRANE PROTEINS APPROACH AND FOLD INTO MEMBRANES

by

Dagan Curtis Hayward Marx

A dissertation submitted to Johns Hopkins University in conformity with the requirements for

the degree of Doctor of Philosophy

Baltimore, Maryland

December 2020

Abstract

Membrane proteins must fold into phospholipid bilayers to function. Some membrane proteins are cotranslationally inserted into the membrane, while others rely on chaperone networks to solubilize and traffic unfolded membrane proteins to the membrane. In my thesis work I have studied both facets of membrane protein folding: chaperone interactions and insertion into the membrane. The first part of my thesis investigates the intrinsic conformational properties and unfolded outer membrane protein (uOMP) binding of the main chaperone in OMP biogenesis pathway in *E. coli*: SurA. We found that SurA is monomeric and exists in three major conformations in solution. Next, we mapped the uOMP binding site on SurA, finding that the least intrinsically populated conformation is the chaperone-active state. Using a plethora of experimental data as restraints, we constructed a model of the SurA-uOMP complex in which the uOMP is greatly expanded by SurA. In the second part of my thesis I examined the effect of the local environment imposed by both the protein and bilayer on individual side chain transfer free energies. I found that the transfer free energies for most side chains are only slightly altered by changes in neighboring side chains and packing. By relating the local composition of the membrane to nonpolar side chain transfer free energies, I discovered a linear correlation between the nonpolar solvation parameter and local concentration of water in the bilayer. Together these studies highlight the structural and thermodynamic parameters that drive efficient membrane protein biogenesis and folding.

Thesis Advisor: Dr. Karen Fleming **Second Reader:** Dr. Stephen Fried

Thesis Committee: Dr. Juliette Lecomte, Dr. Bertrand Garcia-Moreno, & Dr. Jungsan Sohn

Acknowledgements

This thesis is dedicated to my recently departed mother Linda Hayward, PhD., who inspired my love for science and lovingly supported every step I've taken along the way.

In her words:

"It's 'bout damn time!".

List of Abbreviations

ASA: solvent Accessible Surface Area

ATP: Adenosine 5'-triphosphate

BAM: β-Barrel Assembly Machine

CD: Circular Dichroism

DegP: E. *coli* periplasmic endoprotease protein

DLPC: 1,2-dilauroyl-sn-glycero-3-phosphocholine

D_{Max}: Maximum end-to-end Distance

DSBU: Disuccinimidyl Dibutyric Urea

FkpA: E. *coli* FkpB binding protein A

FRET: Förster Resonance Energy Transfer

GdnHCl: Guanidinium Hydrochloride

LPS: Lipopolysaccharide

LUV: Large Unilamellar Vesicle

mOMP: Mitochondrial Outer Membrane Protein

MD: Molecular Dynamics

OMP: Outer Membrane Protein

OmpLA: E. *coli* Outer Membrane Phospholipase A1

PagP: E. *coli* Lipid A Palmitoyl Transferase

TOM: Translocase of the Outer Membrane

NMR: Nuclear Magnetic Resonance

pAF: para-AzidoPhenylalanine

POPC: palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine

PPIase: Peptidyl-prolyl isomerase domain

pXL-MS: Photo-Crosslinking Mass Spectrometry

R_G: Radius of Gyration

RMSD: Root Mean Square Deviation

SAM: Sorting and Assembly Machinery

SANS: Small Angle Neutron Scattering

SDS-PAGE: Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis

SE: Sedimentation Equilibrium Analytical Ultracentrifugation

Skp: E. *coli* Seventeen Kilodaton Protein

smFRET: single molecule FRET

SurA: E. *coli* survival factor protein A

σ_{NP}: Nonpolar Solvation Parameter

uOMP: Unfolded Outer Membrane Protein

UV: UltraViolet

XL-MS: Crosslinking Mass Spectrometry

Table of Contents

List of Tables

List of Figures

Chapter 1 – Introduction to Membrane Protein Biogenesis and Folding

Parts of this chapter are submitted as: Marx, D.C., and Fleming, K.G., (2021) Membranes Proteins Enter the Fold. Current Opinions in Structural Biology.

1.1 General Overview of Membrane Protein Biogenesis

Proteins are biological macromolecules comprised of a linear chain of amino acids that undergo a process called folding to adopt specific three-dimensional conformations (1). There are two main classes of proteins: water-soluble proteins that fold and function in the aqueous cytoplasm and inside organelles, and membrane proteins, which fold into- and function in hydrophobic phospholipid bilayers. Subsequently, there are two groups of membrane proteins that are classified by the basic structural unit formed in their transmembrane domains (TMD): α-helices and β-barrels (2). Membrane proteins perform essential functions such as regulating import and export of molecules to/from the cell, modifying the lipids in the membrane, and creating the cell's main energy source ATP (3–5).

Forming folded structures is crucial for most proteins, which require highly specific three-dimensional atomic arrangements of accurately positioned side chains to perform specific functions such as catalyzing chemical reactions or interacting with other proteins to transmit information through the cell (6). In addition to folding, many proteins require specific localization to sub-regions of the cell such as organelles to perform their functions (7). The folding and trafficking of nascent proteins is combined in the general term "biogenesis" which encompasses intrinsic protein folding, interactions with chaperones and enzymes, and trafficking pathways, among other things (7–9).

For membrane proteins, biogenesis is a complex process. The aggregation propensities of membrane proteins in the aqueous cytosol requires cells to tightly couple membrane protein translation to either folding directly into a membrane or interacting with

chaperone proteins that bind nascent membrane proteins and suppress aggregation (10). The vast majority of membrane proteins contain an N-terminal signal sequence that is recognized by soluble protein chaperones (11). For most alpha-helical membrane proteins, signal sequence recognition triggers the localization of the translation ribosome to the membrane where it interacts with the Sec translocon machinery (12).

The Sec translocon (Sec61 in eukaryotes and SecYEG in prokaryotes) is a membrane protein that acts as a protein-conducting channel and also catalyzes the insertion of transmembrane domains into the bilayer. The prevailing hypothesis for translocon function revolves around a transmembrane "lateral gate" that can open to expose peptides inside the translocon pore to phospholipid bilayer (13, 14). Membrane proteins are thought to partition into the bilayer if the interactions with the bilayer are more favorable than interactions with the translocon (15). An alternative theory considers the channel of the translocon as a mechanism to transport soluble, polar regions of membrane proteins across the bilayer by limiting unfavorable interactions with the bilayer. Transmembrane helices are posited to form in the bilayer interface and inserts into the bilayer if the transmembrane conformation is more energetically favorable than the interfacial conformation (Figure 1.1) (16).

For other membrane proteins, the biogenesis process is more complex and involves post-translational insertion into the bilayer and folding. A subset of alpha-helical membrane proteins called tail-anchored proteins are fully synthesized in the cytoplasm and are shuttled to the ER membrane by a complex chaperone network (17). Mitochondrial outer membrane proteins (mOMPs) are synthesized in the cytosol and

targeted to the mitochondrial outer membrane (18). These nascent beta barrels then cross the mitochondrial outer membrane via the TOM (translocase of the outer membrane) complex to the aqueous intermembrane space inside the mitochondria where they are solubilized by yet another chaperone network (19). Folding of mOMPs is catalyzed by the SAM (sorting and assembly machinery) complex via sequential templating and insertion of beta-hairpin motifs (20). Outer membrane proteins in gram-negative bacteria have a similar biogenesis pathway as mOMPs that is detailed in the upcoming section.

1.2 The *Escherichia coli* Outer Membrane Protein Biogenesis Pathway

Architecture of the Membrane System of Gram-Negative Bacteria

Gram-negative bacteria are defined by the presence of both an inner and outer membrane which are separated by the aqueous periplasm (21). The wide diversity of phospholipids found in the inner membrane generally resembles the plasma membrane of eukaryotic cells (22). The lipid composition of the outer membrane, on the other hand, is distinct from other membranes in biology. The composition of the lipid leaflets of the outer bilayer is asymmetric, with the inner leaflet that faces the periplasm having a similar composition to the periplasmic facing leaflet of the inner membrane. The outer leaflet is comprised of lipopolysaccharide (LPS). LPS contains a variable number of acyl tails (5-8) connected by two sugars that comprise Lipid A (5). Attached to Lipid A are oligosaccharides of varying length that protrude out from the surface of the bacteria.

In addition to lipid composition, the inner and outer membranes also differ in the types of proteins that fold into them (23, 24). The inner membrane contains α-helical membrane proteins, which are co-translationally inserted via the Sec-translocon machinery. These proteins are responsible for many essential cellular processes including maintaining cytoplasmic homeostasis, creating chemical energy for the cell (ATP), and transmitting signals from the extracellular environment and periplasm to the cytoplasm to alter gene expression. The outer membrane, on the other hand, contains proteins with β-barrel structures. These outer membrane proteins (OMPs) vary in size ranging from 8-26 beta strands, with larger OMPs seemingly requiring small, folded plug domains to stabilize the beta-barrel structure. Many OMPs are thought to be monomeric

when folded in the outer membrane, however some OMPs can also exist in higher-order oligomeric states (25). Some OMPs form obligate trimers (such as OmpC and OmpF), and others (such as OmpLA) couple their oligomeric state (and enzymatic function) to chemical stimuli such as calcium (26–28).

The inner and outer membranes of gram-negative bacteria are separated by the aqueous periplasmic space which contains a variety of proteins and the peptidoglycan layer (24, 29, 30). The peptidoglycan layer in gram-negative bacteria is thought to be relatively thin, with pores of varying sizes and is covalently linked to the outer membrane (31). Together, the peptidoglycan and outer membrane create the load-bearing element for gram-negative bacteria (32, 33). The soluble proteins in the periplasm perform a variety of functions such as trafficking proteins and lipids across the periplasmic space to the outer membrane, creating and altering the structure of the peptidoglycan layer, and sensing environmental changes (34).

Outer membrane protein biogenesis in E. coli

To fold into the outer membrane, OMPs are translated in the cytosol and must cross both the inner membrane and the periplasm. The resulting journey from the ribosome to the outer membrane is termed the OMP biogenesis pathway and is illustrated in Figure 1.2 (24). The first step in this biogenesis pathway is post-translational secretion across the inner membrane via the Sec translocon. OMPs are synthesized with an N-terminal signal sequence, which directs the cell to traffic the nascent OMP to the inner membrane. The cytoplasmic chaperone SecB is thought to recognize and bind to most OMP as they are translated in the cytoplasm and brings them to the SecYEG translocon

(35, 36). Secretion of OMPs through the translocon is dependent on SecA, which couples ATP hydrolysis to pushing the nascent, unfolded OMP (uOMP) through the translocon (37). Secreted OMPs are tethered to the inner membrane until the signal sequence is cleaved by signal peptidase, at which time they are fully released into the aqueous periplasm (38).

In the aqueous periplasm uOMPs encounter a network of chaperones that solubilize uOMPs to prevent aggregation and other off-pathway conformations. This network is comprised of three main chaperones: survival factor A (SurA), FK-binding protein A (FkpA), and seventeen kilodalton protein (Skp) (24, 39, 40). The periplasmic serine endoprotease DegP is also part of the network and is thought to sense the amount of uOMP (41). Together, this network binds uOMPs in the periplasm through distinct mechanisms and somehow directionally traffics uOMPs across the peptidoglycan layer to the outer membrane. At the outer membrane, uOMPs encounter the beta-barrel assembly machine (BAM) complex which catalyzes the folding of uOMPs into the outer membrane of E. *coli* (24, 42, 43).

The mechanism by which the members of the uOMP biogenesis chaperone network act in concert to ferry uOMPs across the periplasm is still unclear. It has been recently shown for the LPS biogenesis pathway that a transient protein bridge spanning the periplasm is formed that allows for efficient transfer of these molecules (44). This bridge requires the machinery for LPS export found in the inner membrane to be localized with the OMP LptD which inserts LPS into the outer membrane. Recently a similar bridge

has been described for OMP biogenesis which would allow for uOMPs to be almost directly handed off from the Sec translocon to the BAM complex (45).

However, this finding is somewhat incongruous with the experimentally measured two minute lifetime of uOMPs in the periplasm (46). Using known binding kinetics and thermodynamics data, the uOMP chaperone network was modeled using a series of differential equations (47). This study revealed that solely modeling the intrinsic interactions between chaperones and uOMPs could recreate both the uOMP periplasmic lifetime and the copy number of uOMPs folded into the outer membrane. The structural basis for the uOMP biogenesis pathway remains unclear, as do the varying roles of each of the periplasmic chaperones SurA, FkpA, and Skp. However, one common feature between the two proposed models of uOMP biogenesis is the feature of SurA as an important player in the pathway leading to the efficient folding of OMPs.

1.3 The Role of SurA in the OMP Biogenesis Pathway

Overview of the role of SurA in the periplasm

SurA is thought to be the "most important" chaperone in the uOMP biogenesis pathway as it is the only chaperone that produces a phenotype when it is knocked out (Δ*surA*) (34, 48–52). While SurA is not essential for E. *coli* proliferation, tandem deletions of SurA and either Skp or DegP induce a more severe phenotype than Δ*surA* alone (53, 54). This finding implies that there are overlapping or compensatory functions for Skp and DegP that are exacerbated when SurA is absent. It is likely that SurA is the main chaperone in the pathway and handles the majority of uOMP flux while Skp and FkpA potentially act as a backup reservoir of uOMP binding capability. As DegP functions to degrade uOMPs not bound to chaperones, its function is most likely critical in the absence of SurA to degrade uOMP to avoid aggregation (41).

SurA is thought to have a preferred subset of eight OMPs (out of 23 total OMPs in E. *coli*) that it chaperones *in vivo*, termed clients (39, 55, 56). These client OMPs include the essential LPS biogenesis protein LptD, proteins involved in the transport of iron (FecA and FhuA), sugar (LamB), and fatty acid (FadL) molecules across the outer membrane, and presumed structural proteins OmpA, OmpX, and OmpF. Studies of binding energetics of SurA to client and nonclient uOMPs have found dissociation constants in the low micromolar range (1-10 μM), which indicates that binding in the periplasm may be non-stoichiometric given the estimated concentration of SurA is only 10-20 μM (55, 57–64).

In addition to acting as a chaperone for its client uOMPs, there has been an increasing amount of genetic and *in vitro* evidence that SurA interacts with the folded BAM complex. The Silhavy group has found that mutations to proteins in the BAM complex that induce phenotypes in E. *coli* can be rescued by compensatory mutations in SurA (40, 65). This genetic interaction of SurA and BAM has hinted that SurA may be directly involved in the delivery of uOMPs to the BAM complex. Additionally, crosslinking mass spectrometry experiments have corroborated this interaction, with multiple crosslinks found between SurA and BamA (66).

It is becoming clear that SurA may function as an inner-to-outer membrane liaison for OMPs, as multiple studies have shown evidence for SurA to also interact with a Sec translocon-based complex in the inner membrane. SurA was found to crosslink to PpiD, a proline cys-trans isomerase which is known to interact with the SecYEG translocon (66). Paired with the interaction with the BAM complex, it is enticing to view SurA as the molecule the bridges the inner and outer membrane machinery responsible for OMP biogenesis. Recent cryo-EM studies have found that a complex can form directly linking the BAM complex with the Sec translocon complex (SecYEG, A, DF) and YidC with and without SurA (45). This super-complex also implicates SurA in aiding the directed transport of uOMPs across the periplasm. Structural studies also found evidence for SurA to interact directly with the translocon, indicating that in some instances SurA could directly recognize and solubilize uOMPs as they are secreted into the periplasm .

Together, the identification of two different complexes linking SurA to both the BAM complex and the Sec translocon highlight the relative importance of SurA compared to

the other chaperones in the periplasmic uOMP biogenesis pathway. It remains to be seen whether these periplasm spanning complexes are the norm for uOMP biogenesis or if they exist only under certain circumstances such as stress conditions or periods of fast growth where streamlining OMP production may be beneficial to the cell. Additionally the structural details of these interactions between SurA and both the outer and inner membrane components of the OMP biogenesis pathway need to be elucidated in greater detail to form concrete hypotheses regarding these various mechanisms for SurA to chaperone uOMPs in the periplasm.

Structure and Dynamics of SurA and Implications on Chaperone Function

The monomeric crystal structure of full-length SurA was determined at 3.0 Å resolution and shows that SurA has three structural domains: a core chaperone domain and two peptidyl prolyl isomerase (PPIase; P1 and P2) domains (Figure 1.3A) (67). The core domain is comprised of a complex formed by the N- and C-terminal regions of the protein and has been shown to compliment the Δ*surA* strain, indicating that the core domain may be responsible for chaperone function. The P1 and P2 domains are connected to each other, and to the core domain by flexible linkers that are proposed to allow for SurA to have a conformationally dynamic structure in solution. Indeed, SurA is structurally distinct from the other periplasmic members of the uOMP biogenesis chaperone network. SurA is proposed to function as a monomer in solution while the other chaperones in the network, Skp (trimeric), FkpA (dimeric), and DegP (multiple proposed higher order stoichiometries), are proposed to oligomerize in order to bind and solubilize uOMPs (24, 47).

A standing question in the field centers around the roles of the P1 and P2 domains, particularly their conformational dynamics, in SurA chaperone function. In the monomeric crystal structure of SurA, the core and P1 domains form a complex while the P2 domain is structurally isolated from the other domains (Figure 1.3C). A dimeric crystal structure of SurA was also solved using a construct lacking the P2 domain (SurAΔP2) in complex with a small peptide (Figure 1.3B) (68). In this structure, the core-P1 interface of each monomeric is abolished and the P1 domain is extended away from the core domain (Figure 1.3D). However, there are extensive core-core and P1-P1 contacts in each monomer in this structure, so the formation of this core-P1 conformation in solution as a monomer was still in question.

These two structures hint at possible dynamics of the PPIase domains of SurA. Accordingly, both P1 and P2 domains' roles in binding uOMPs have also been investigated. SurAΔP2 was found to have no effect on the binding affinity to the client OmpA, but drastically decreased binding to the nonclient OmpT (57). Binding to OmpA was slightly weakened when both PPIase domains were deleted, indicating that the core domain is not solely responsible for uOMP binding. Together, these results indicate that the structure of SurA may not be static in solution and that core and P1 domains are essential for recognizing and binding client uOMPs while P2 may serve to discriminate between client and nonclient uOMPs. However, this study did not investigate the binding affinity of a construct lacking either the P1 or core domains, which would be required to completely understand the thermodynamics driving SurA-client uOMP interactions.

Subsequently, the dynamics of the domains in SurA have been implicated in regulating chaperone function through both *in vitro* and *in vivo* experiments. Mutations to the core-P1 interface have been shown to alter the apparent chaperone function of SurA *in vivo*. Removing a serine (S220A) in the P1 domain located at the core-P1 interface was shown to rescue a deleterious mutation to the BAM complex, indicating that the activity of SurA and the BAM complex are correlated (40). To further explore the importance of the core-P1 domain interaction, cysteines were introduced on both the core and P1 domains that would covalently link the core-P1 complex closed via disulfide bonds. The cysteine variants of SurA were found to impair chaperone function, indicating that disruption of the core-P1 complex is potentially a requirement of chaperone function.

Several recent studies, including Chapters 2 and 3 of this thesis, have investigated the conformations of the SurA domains *in vitro* using a combination of ensemble and single-molecule techniques (60, 69, 70). Single-molecule fluorescent resonance energy transfer (FRET) experiments showed that the P2 domain can exist closer to the core-P1 complex in solution than is shown in the crystal structure. This finding was corroborated by both NMR and crosslinking mass spectrometry experiments. However, the functional implications of the spatial relationship between the core-P1 complex and the P2 domain are still unclear. The gain-of-function S220A variant was also studied using NMR and single molecule FRET, which corroborated the interpretation of the *in vivo* findings that the mutation does induce disassociation of the core and P1 domains. Additionally, the introduction of an OMP-mimicking peptide induces a similar shift in the orientations of the

core and P1 domains NMR studies also detected some interaction between the core and P2 domains, which was observed in molecular dynamics simulations of SurA as well.

Questions Remain Regarding the Chaperone Function of SurA

While the structure and importance of the periplasmic chaperone SurA have been known for over a decade, the mechanism by which SurA performs its functions is unknown. SurA is assumed to be monomeric in solution primarily due to its crystal structure, however this has not been experimentally assessed. Additionally the conformational ensemble of SurA is not yet established quantitatively. Recently published fluorescence and NMR experiments establish the presence of multiple conformers, but do not define the population weighted conformational ensemble.

Aside from the intrinsic structure of SurA in solution, the chaperone active conformation of SurA is not known. The disordered and aggregation-prone nature of uOMPs has hampered structural characterization of the SurA-uOMP complex. Most structural techniques require high concentrations of protein, which is untenable for unfolded membrane proteins. The stoichiometry of the SurA-uOMP complex is apparently complicated, with binding titrations revealing an m-value greater than one, but less than two (57). How this intermediate stoichiometry is realized in solution is unclear, but it points to the ability of more than one SurA to bind to a single client uOMP.

On top of the questions about SurA when bound to uOMPs, it is unknown which regions of uOMPs interact with SurA. A single study using peptide arrays implicated aromatic residues in a proposed SurA-binding motif, however the low resolution nature of the study approach has left the actual regions of uOMPs that SurA unclear (55). This

study identified more than one segment of uOMP sequence that bound SurA, which perhaps hints at the manifestation of the non-straightforward stoichiometry of the SurA-uOMP complex.

Finally, the mechanism for solubilization of uOMPs by SurA has yet to be established. Skp, one of the other chaperones in the uOMP biogenesis network, trimerizes upon binding to sequester uOMPs from aqueous solution to avoid aggregation and misfolding events (71, 72). FkpA is dimeric and contains a large cavity between monomers that is hypothesized to function similarly to Skp. SurA does not have an obvious ability to form a cage-like structure based on the structural findings to this point. Does SurA form a cage-like structure by having multiple monomers bind a uOMP at a given time? Or does SurA solubilize uOMPs though an alternate mechanism than the other periplasmic chaperones? Defining the structure and dynamics of apo- and uOMP-bound-SurA is critical for understanding, and eventually targeting, the essential uOMP biogenesis pathway of gram-negative bacteria.

1.4 Overview of Membrane Protein Folding and Stability

The Value of Water-to-Bilayer End Points

As discussed previously, membrane proteins must fold into phospholipid bilayers to perform their cellular functions. The prevailing, general model for membrane protein folding is broken down into two steps (73). The first step involves the insertion of individual transmembrane domains into the membrane and the second involves interactions between TMDs to form stable tertiary and quaternary structures. Overall, the process of membrane protein folding can be described by the three equilibria shown in Figure 1.4.

Water-solvated unfolded, U_W, and bilayer-embedded folded states, F, represent the two most extreme endpoints of biophysical interest for membrane-protein folding reactions. A deceptively simple parameter – the free energy of folding ($\Delta G^0_{U_W,F}$) – captures the population bias at equilibrium, and the free energy change between these end states reveals the maximum energetic contributions of the various atomic interactions responsible for stabilizing a particular folded state over its aqueous-unfolded conformational ensemble. Although the water-soluble unfolded state is not typically observed in a cellular context, these endpoints are nevertheless useful in theoretical considerations that seek to describe the underlying chemical reactions.

Taking cues from the soluble protein-folding field, a number of groups used chemical denaturation titrations and extensive condition tweaking to measure path-independent equilibrium values for several transmembrane β-barrels (74–76). These experiments reveal an extremely favorable folding stability for β-barrels, ranging from -18

to -32 kcal mol^{-1}, and the systems have proved useful in addressing the energetic contributions of side-chain partitioning and backbone hydrogen bond formation (75, 77–80).

Membrane-embedded Unfolded-to-Folded Endpoints Dominate α-Helical Membrane Protein Measurements

To date, there are no water-to-bilayer stabilities measured for α-helical transmembrane proteins. This is presumably due to the enhanced aggregation propensities of transmembrane α-helical regions that are comprised of continuous stretches of nonpolar amino acids. Stability measurements of α-helical membrane proteins have accordingly been tractable only in experimental setups in which unfolded states remain embedded in a membrane or in a membrane mimic, which we term U_M, regardless of its secondary structure. In these reactions the energy derived from the hydrophobic effect is attenuated because the water concentration is not bulk, and a smaller free energy difference between U_M and F is expected. If the α-helical secondary structure is stable in isolated segments in the unfolded ensemble, e.g. $U_{M,\,H}$, these experiments should report on transmembrane helix-helix interactions, e.g. $U_{M,\,H} \leftrightarrow F$.

The classic example of this reaction includes the dimerization of the single-transmembrane domain of glycophorin A, GpATM (81–83). However, new methods that interrogate helix-helix interactions in more complex multi-span proteins show that the lateral interactions are not going to be simple to understand. Local interactions show varied stabilities in the intramembrane rhomboid protease GlpG as assessed using a 'steric trapping' strategy (84–87).

In contrast, the ClC-ec1 Cl^-/H^+ antiporter has a high affinity in bilayers using a promising new single-molecule microscopy technique (88–90). The method is model-independent and can be carried out in any bilayer of choice using single-molecule fluorescence bleaching steps to quantify the membrane protein oligomer size following equilibration in what is essentially an "infinite" bilayer. In 2:1 POPE:POPG, the authors found that ClC-ec1 forms a high-affinity dimer with a mole fraction equilibrium dissociation constant equal to 4.7 $x10^{-8}$ subunits $lipid^{-1}$. For context, this is only ~1.3 kcal mol^{-1} less favorable than the GpATM dimer in POPC (83), which was a surprising outcome because the ClC-ec1 dimerization interface is much larger by comparison. Because the CLC-ec1 lacks a so-called GxxxG dimerization motif, future mutational analysis on this protein will be needed to rationalize the distinct physical mechanisms these two proteins employ in subunit recognition. The distinction between these two structural modes for dimerization also begs the question of whether the packing of nonpolar side chains is sufficient to drive protein-protein interactions in lipids, which is an area of high interest in the membrane protein design field (91).

Energetic Features of Native Folds

Since the availability of the earliest membrane-protein structures, it has been observed that most membrane proteins are enriched in either transmembrane α-helical or β-sheet (barrel) secondary structure that is formed by regular patterns of backbone hydrogen bonds. Backbone hydrogen bond (bbHB) formation is favored in membrane-embedded regions because there is a larger energetic penalty for the water-to-bilayer partitioning of the non-hydrogen bonded backbone. Recent advances in NMR

experimental methodologies have allowed for bbHB strengths to be measured both in α-helical and β-barrel transmembrane proteins using hydrogen-deuterium exchange (92). Cao *et al.* reported that bbHB strengths for the transmembrane α-helical amyloid precursor protein reach -6 kcal mol^{-1}, a value much more favorable than previous estimates using organic solvents and small peptides, or even soluble proteins (80, 92). Lessen *et al.* performed similar experiments using the transmembrane β-barrel OmpW and found strengths ranging from -3 to -4 kcal mol^{-1} on average (79). In contrast to the partitioning free energy changes of nonpolar side chains discussed above, both NMR investigations found bbHB strengths to be relatively insensitive to the position of the membrane. Together, these studies indicate that bbHB energies appear to be affected by neither sequence nor secondary structure. In sum, the unchanging bbHB energy in membrane proteins across the bilayer implicates side-chain partitioning interactions as the main driving force for transmembrane protein insertion into the bilayer.

Side-chain entropy can be another energy source in protein folding. Compared to U_w, in which the polypeptide chain can assume a large and heterogeneous conformational ensemble, the folding of a transmembrane α-helix upon insertion limits the conformational space and perhaps the motions of side chains (93). In contrast to this assumption, solution NMR relaxation studies suggest membrane proteins are extraordinarily dynamic with fast internal motions on methyl-bearing side chains (94). This finding was equally true for the α-helical sensory rhodopsin II as well as the OmpW β-barrel and was independent of the hydrophobic, membrane-mimicking cosolvent. The

energetic contribution of side-chain motion to folding will depend on the extent to which it is preferentially enhanced in F as compared U_w. Crucially, this remains to be tested (94).

1.5 Side Chain Transfer Free Energies Dominate Membrane Protein Stability

Contribution of Side Chains to Water-to-Bilayer Membrane Protein Folding

In addition to side chain entropy and backbone hydrogen bonds, side chain transfer free energies contribute to the water-to-bilayer membrane protein folding. This energy quantifies the preference of a given side chain to partition into the bilayer or bulk water. Historically, the partitioning of peptide segments between water and a hydrophobic organic solvent has been employed to mimic this energetic contribution as manifested through the construct of a hydrophobicity scale (95, 96). These hydrophobicity scales rank the affinity for each side chain for interacting with the center of the membrane. Generally these scales report similar findings, with nonpolar side chains preferring the organic solvent, and polar residues preferring water.

Moving the membrane mimic from an organic solvent to an actual bilayer brought an understanding of hydrophobicity closer to the cellular condition. Moon and Fleming developed a method to reversibly fold and unfold β-barrel OMPs into DLPC vesicles. Using guanidinium hydrochloride titrations at low pH (3.8) to protonate Glu and Asp residues with ~2 days of equilibration time, the equilibrium population of folded and unfolded OMP can be measured by intrinsic tryptophan fluorescence (97, 98). In total, three OMPs: OmpLA, OmpW, and PagP were found to fold reversibly (76).

Using this technology, Moon and Fleming used a host-guest approach to create a hydrophobicity scale using a whole protein folded inside a phospholipid bilayer (75). In this original work, they assessed side chain hydrophobicity at the center of the bilayer using OmpLA by mutating a host side chain (A210) to every other side chain (guest) and

measuring the thermodynamic stability of each variant. By taking the difference in the stabilities of the host and guest variants of OmpLA, the contribution of the guest side chain to folding can be isolated. This water-to-bilayer hydrophobicity scale is similar to the water-to-organic solvent scales, indicating that these measurements were capturing the chemical nature of the center of a lipid bilayer.

Local Environment Affects Side Chain Transfer Free Energies

Hydrophobicity scales are most widely used to detect potential TMDs from amino acid sequence because side chain transfer free energies are what discriminates between soluble and membrane-embedded structures. In theory, side chain transfer free energies can also be used to accurately estimate the stability of a TMD in the bilayer. To accurately estimate TMD stability it is important to understand what factors can modulate the magnitude or sign of a side chain transfer free energy. In general there are two ways in which the local chemical environment around a side chain in a TMD can be altered: changing the protein sequence/structure and changing position in the membrane.

The membrane itself is not a uniform solvent; rather, amphipathic nature of phospholipids creates a steep polarity gradient that separates the bilayer center from bulk water. This region, termed the bilayer interface, is a chemically complex environment with a steeply changing water concentration. Initially, the hydrophobicity of the bilayer interface was measured by quantifying the water-to-interface partitioning of peptides and a POPC bilayer (99). This study found that side chain transfer free energies were much more muted in the interface than the center of the bilayer, confirming the bilayer position-dependence on side chain transfer free energies.

Using the whole protein, water-to-bilayer approach, Moon and Fleming measured the bilayer position dependence of Arg and Leu by applying the host-guest methodology at other sites on OmpLA (75). They found that the transfer free energy for Leu decreased and Arg increased moving from the center to the edge of bilayer. McDonald and Fleming applied this same approach to measuring the bilayer position-dependence of aromatic residues, finding a preference for the bilayer interface over the hydrophobic core. Importantly, this work incorporated all-atom molecular dynamics simulations to measure the exact bilayer position of each side chain and couple that position to bilayer chemical composition. This is the first report showing the coupling of local bilayer chemical composition with experimentally measured side chain transfer free energies. By understanding how local water or polar atom concentration modulates side chain transfer free energies, these thermodynamic values can be generalized to other proteins and bilayers.

Protein structure and sequence can also modulate the apparent transfer free energy for a given side chain. Moon and Fleming measured the effect of incorporating two Arg residues in the same TMD, finding the energetic penalty is less than the sum of the individual transfer free energies of each Arg independently (75). McDonald and Fleming observed aromatic-aromatic interactions could increase the apparent transfer free energy by 1-2 kcal mol^{-1} (78). While specific instances of local sequence and structure affecting side chain transfer free energies have been observed, no study has addressed this point directly. The local organization of side chains varies based on secondary structure type, and in the case of β-barrels, the number of strands in the

barrel. Additionally, the orientation or tilt of a TMD in the bilayer varies drastically between proteins. Studies measuring the position-dependence of side chain transfer free energies in α-helices have used helical position as a proxy for bilayer position, which generally ignores these protein-to-protein variations in structure and sequence.

Questions about Side Chain Transfer Free Energies Going Forward

Statistically, membrane-embedded segments are highly enriched in apolar side chains that favorably interact with the nonpolar core of the bilayer.(100) One key question concerns how much energy is gained by the removal of a nonpolar moiety from water and its placement within the bilayer. How does this aqueous gradient change the driving force energy of the hydrophobic effect along the bilayer normal? Historically this has been quantified using the nonpolar solvation parameter, σ_{NP}, which quantifies the transfer free energy of 1 Å^2 of nonpolar surface area. Understanding the bilayer position, and crucially the local chemical environment, dependence of the σ_{NP} will allow for the value of nonpolar side chain transfer free energies to be easily generalizable.

The generalizability of hydrophobicity scales has always been assumed and never tested explicitly. How do side chain transfer free energies change when measured in two different proteins under the same experimental conditions? Do nearest-neighbor side chains have a large effect on the apparent hydrophobicity of a given side chain? Does either the packing between side chains or TMD tilt in the membrane modulate transfer free energies in a significant way? Answering these questions is critical for increasing the resolution and applications of side chain transfer free energies for predicting TMD stability and structure.

1.6 Outline of Data Presented Here

Membrane proteins must fold into phospholipid bilayers after being synthesized in an aqueous environment. The process by which membrane proteins approach and fold into the membrane is essential for cell viability across all domains of life. Despite the importance of membrane protein biogenesis, fundamental questions remain regarding both steps of this process. In my thesis I have interrogated structural and energetic determinants of membrane protein biogenesis through a combination of experiment and computation.

A membrane protein biogenesis pathway of particular biomedical importance is the E. coli OMP biogenesis pathway. This pathway has been shown to be a promising target for antibiotic drug development because drugs would not have to cross the tightly regulated inner membrane, but the majority of the steps of this pathway are structurally and functionally unresolved. I have studied the periplasmic chaperone SurA, which is thought to be the most important chaperone in the pathway and is expected to facilitate the penultimate step by delivering OMPs to the BAM complex.

In Chapter 2 I present my investigations on the intrinsic conformational and oligomeric properties of SurA in the absence of membrane proteins. I discovered that SurA is monomeric at physiologically relevant conditions, which distinguishes it from the other chaperones in the OMP biogenesis pathway which form higher-order, oligomeric structures. In addition to the oligomeric state of SurA, I also investigated the conformational dynamics of the three domains of the SurA monomer in solution. I found that the average SurA structure in solution contains a complex between the core and P1

domains, with the P2 domain as a structurally isolated, satellite domain. While this may be the average SurA conformation, I measured the binding affinity of both the core-P1 and core-P2 interactions and found that the two PPIase domains compete for binding to the core domain.

Building on this knowledge of the intrinsic properties of apo-SurA, in Chapter 3 I describe the structural determinants of how SurA binds and solubilizes unfolded OMPs. I found that the chaperone-active conformation of SurA is the "open" conformation in which each domain of SurA is disassociated from one another and that uOMPs bind in a groove found between the core and P1 domains of SurA. SurA is able to recognize and bind to multiple segments of client uOMPs. Surprisingly, the size of uOMPs increases almost two-fold when bound to SurA. The expansion of uOMPs combined with the specific SurA-binding regions has multiple implications for the OMP biogenesis pathway.

In regards to the folding and stability of membrane proteins, I have also sought to better understand how the local chemical environment defined by both protein and bilayer composition affect side chain transfer free energies. In Chapter 4 I measured a hydrophobicity scale for all twenty naturally occurring amino acids at the center of the bilayer using the *E. coli* OMP PagP as the protein scaffold. This study, combined with the analogous study from Moon and Fleming which used OmpLA as the protein scaffold, allowed for the effects of protein structure and sequence on side chain transfer free energies to be interrogated for the first time. I found that the protein scaffold in which a side chain exists has a limited effect on the magnitude of the side chain transfer free

energy the majority of the time. This finding allows for increased confidence of applying side chain transfer free energies broadly to any membrane protein.

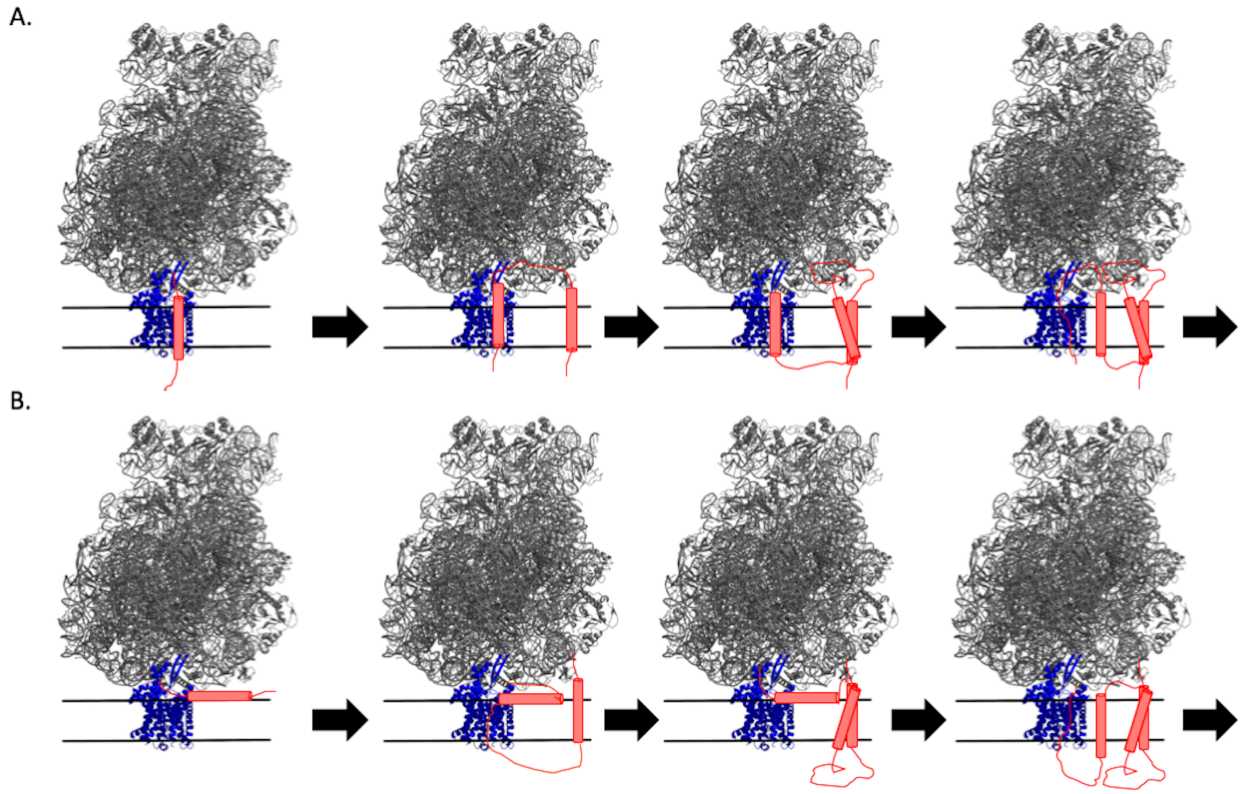
In Chapter 5 I investigated how the magnitude of nonpolar side chain transfer free energies is modulated by the local chemical composition of the bilayer in which the side chain resides in the folded protein. I was able to derive the nonpolar solvation parameter, which correlates transfer free energies with the amount of nonpolar surface area of the side chain, as a function of position in the bilayer. This solvation parameter was found to have a linear correlation with the local concentration of water in the bilayer, allowing for the parameter to be calculated at any position in a bilayer (assuming the water concentration gradient has been determined). Comparing the data collected in this study measuring water-to-bilayer side chain transfer free energies with previously published translocon-to-bilayer transfer free energies I was able to determine that the translocon generally mimics the bilayer interface energetically. This finding means that the translocon-to-bilayer transition is equivalent to the interface-to-bilayer transition, which supports a model for translocon-mediated membrane protein insertion proposed by Cymer, von Heijne, and White (16).

Together, these investigations reveal how interactions with both chaperone proteins and the membrane itself can modulate membrane protein structure and stability. The novel insight into the OMP biogenesis pathway will hopefully allow for antibiotic drugs to be designed to inhibit SurA chaperone function. Additionally, these findings provide multiple avenues of future investigation of the role and function of SurA in forming heterocomplexes with other chaperones in the E. coli periplasm. The transfer free energy

measurements should increase the accuracy of computational algorithms that seek to predict transmembrane structures and estimate membrane protein stability from sequence data. Additionally, the derived correlation between bilayer position and side chain transfer free energies allows for the difference in the stability of different membrane protein conformers to be estimated and incorporated in simulation forcefields.

1.7 Figures

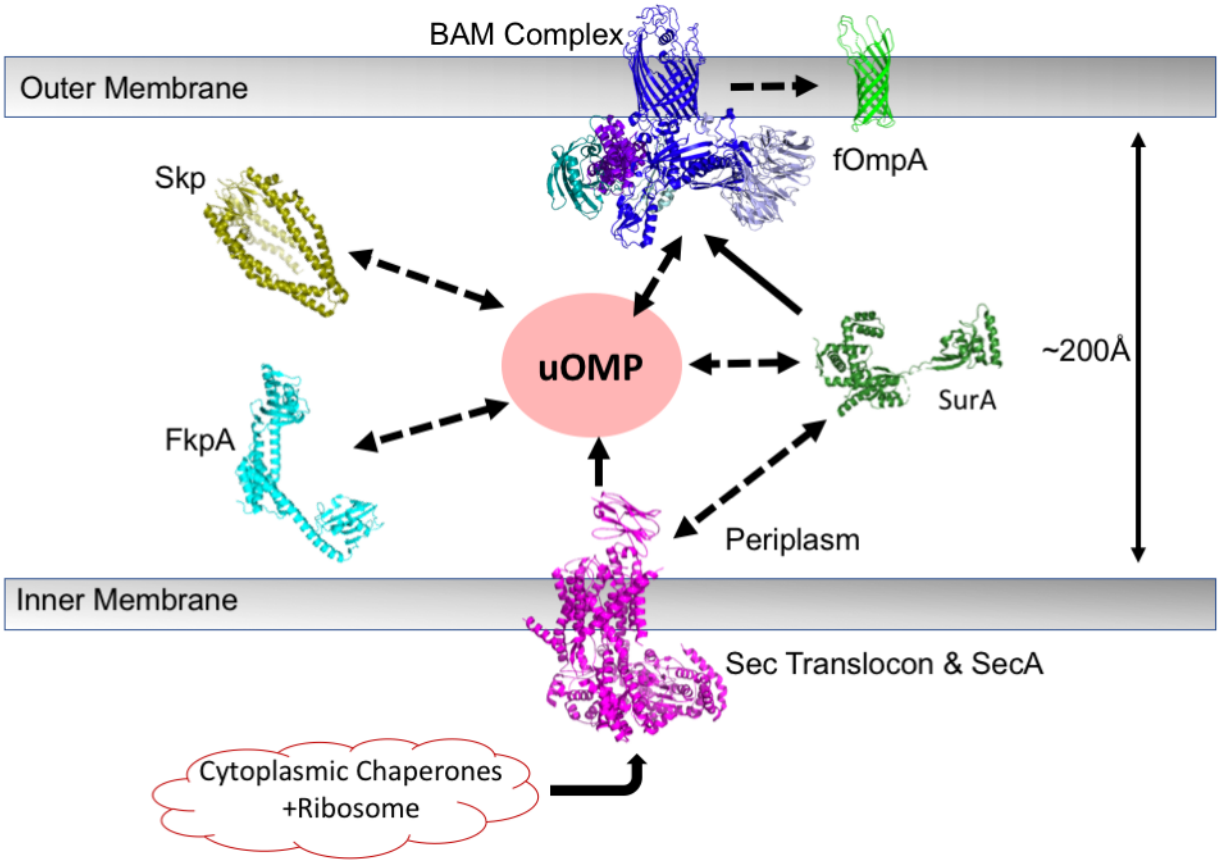
Figure 1.1 – Models of Translocon-Assisted Membrane Protein Folding



Most α-helical membrane proteins are co-translationally inserted into the membrane using the Sec translocon. The *E. coli* SecYEG is colored blue, the ribosome is colored gray, the membrane is shown as black spheres placed by the Orientations of Proteins in Membranes server, and the translating membrane protein is depicted as a red cartoon (PDB: 5GAE) (101, 102). Two models for translocon-mediated TMD insertion have been proposed: **(A)** the translocon serves as a channel with a lateral gate that allows TMD to partition into the bilayer, and **(B)** the translocon serves as a pore for hydrophilic segments of the protein to cross the membrane while TMD form in the bilayer interface and insert into the membrane based on the thermodynamics of the interface-to-bilayer transition for that TMD. Both mechanisms result in the same folded conformation of the membrane

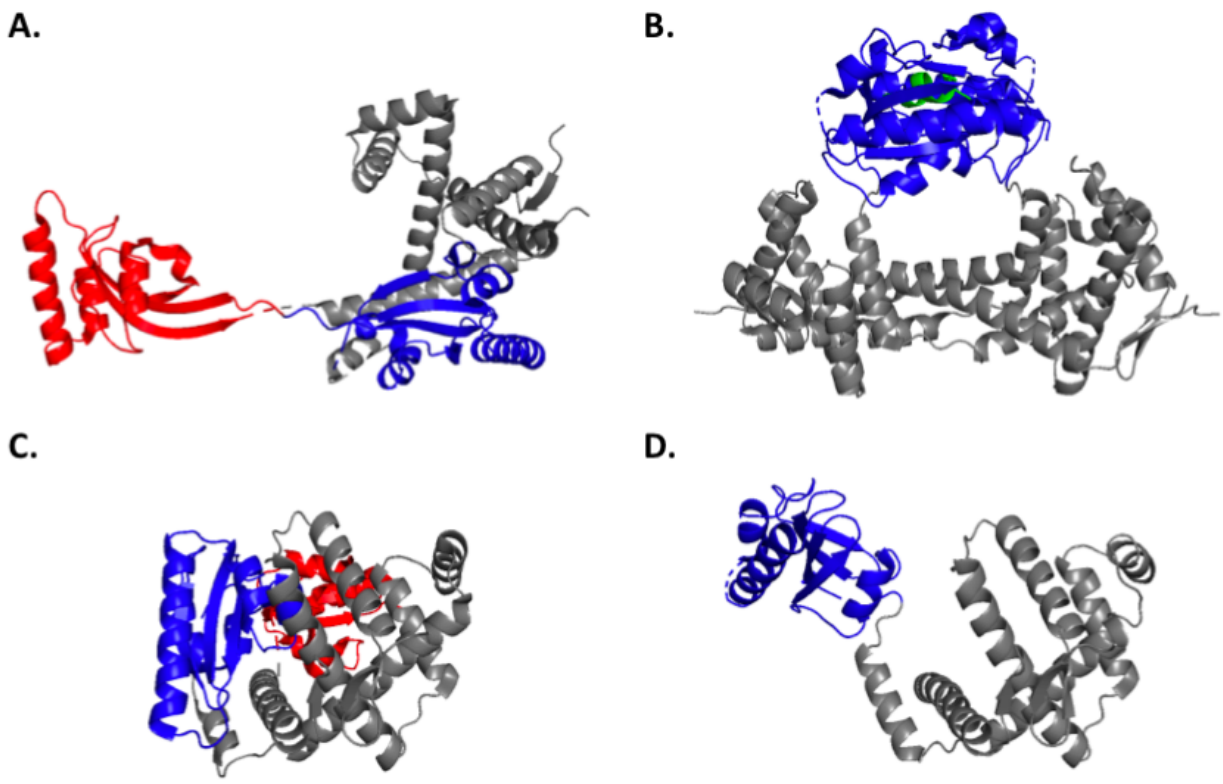
protein and differ based on the role of the translocon in assisting TMD to insert in the bilayer.

Figure 1.2 The *E. coli* OMP Biogenesis Pathway



OMP biogenesis is a multi-step process, ending with folding into the outer membrane in E. *coli*. To reach the outer membrane uOMPs are translated in the cytoplasm and are solubilized by cytoplasmic chaperones. These chaperones deliver nascent uOMPs to the SecYEG-SecA complex (magenta, PDB: 5EUL). Upon translocation across the inner membrane, uOMPs enter the periplasm (red oval) where they can interact with the periplasmic chaperone network comprised of SurA (dark green, PDB: 1M5Y), Skp (gold, PDB: 1SG2), and FkpA (cyan: 1Q6U). To fold into the outer membrane, uOMPs must interact with the BAM complex which is comprised of BamA-E (various shades of blue, PDB: 5AYW). SurA is thought to hand-off uOMPs to the BAM complex (solid arrow), which catalyzes folding of uOMPs into the outer membrane (light green, PDB: 1BXW).

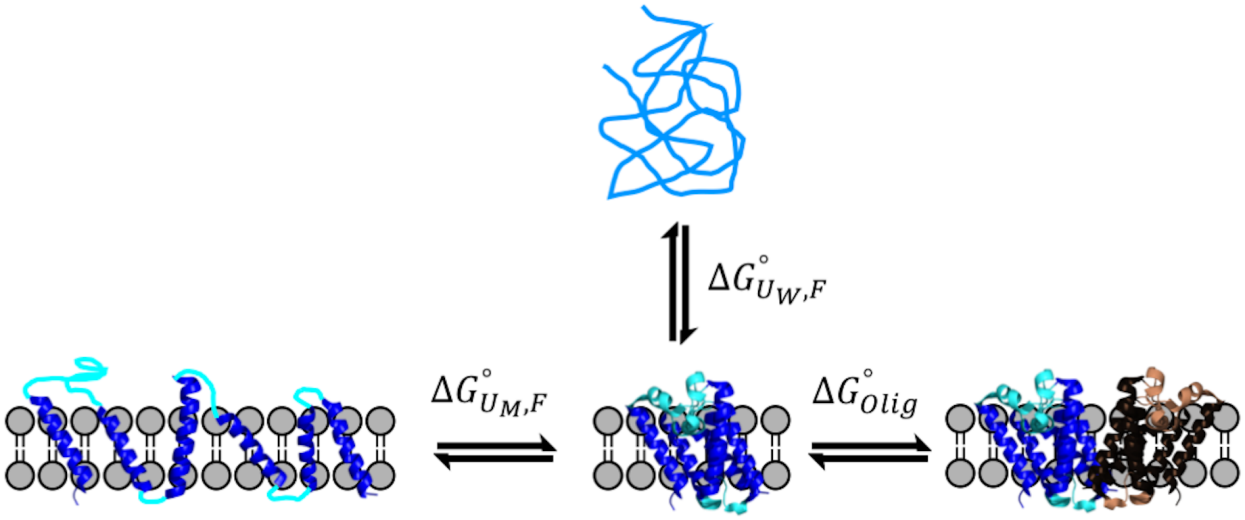
Figure 1.3 - Crystal Structures of SurA Hint At Conformational Dynamics



(A) The monomeric crystal structure of SurA (PDB: 1M5Y) is in a cartoon representation. The core (gray) and P1 (blue) domains are interacting and the P2 domain (red) is structurally isolated from the core-P1 complex. **(B)** The dimeric crystal structure (PDB: 2PV3) of SurAΔP2 is shown with a cartoon representation. SurAΔP2 was crystallized in complex with a uOMP mimicking peptide (green helix) that is bound to the P1 domains of the dimer. In this structure, core-core and P1-P1 complexes are formed with the core-P1 interaction seen in **(A)** abolished. **(C)** Monomeric SurA is shown rotated to have the P2 domain situated behind the core-P1 complex. **(D)** One of the monomers from **(B)** is shown with the core domain aligned to the core domain in **(C)**. The structures in Panels **C** and **D** highlight the conformational change of the P1 domain in the two crystal structures of SurA.

The ability of the core-P1 unit to adopt different conformations in solution suggests that SurA may adopt multiple conformations.

Figure 1.4 – Membrane Protein Folding Equilibria



The relevant thermodynamic equilibria describing membrane protein stability and the experimental approaches used to measure each free energy are shown. The protease GlpG (PDB: 3B45) is shown as a cartoon representation to illustrate each transition. $\Delta G^{\circ}_{U_W,F}$ describes the coupled folding and insertion of an unfolded, water-soluble membrane protein into the bilayer, $\Delta G^{\circ}_{U_M,F}$ describes the association/folding of helices in a membrane unfolded state, and ΔG°_{Olig} describes the oligomerization of membrane proteins.

Chapter 2 – Domain Interactions Determine the Conformational Ensemble of the Periplasmic Chaperone SurA

Published as: Marx, D.C., Leblanc, M.L., Plummer, A.M., Krueger, S., and Fleming, K.G., (2020) Protein Science, 29(10)2043-2053

2.1 Introduction

Outer membrane proteins (OMPs) play essential roles in gram-negative bacteria such as nutrient uptake (103), modifying lipid structures (104, 105), and rigidifying the cell (32, 33). OMPs are post-translationally secreted into and subsequently trafficked through the aqueous periplasm by a network of chaperone proteins that bind to unfolded OMPs (uOMPs) to prevent aggregation and misfolding (24, 47, 50). Folding is finally catalyzed by the β-barrel assembly machine (BAM) at the outer membrane (42, 106, 107). SurA is thought to be the most important chaperone in the OMP biogenesis pathway as it has been shown to interact with a subset of OMPs (51) (termed "clients") at every stage of this pathway: it is associated with OMPs at the translocon during secretion into the periplasm (45), in the aqueous periplasm (56), and is thought to hand-off clients to the BAM complex (66).

Structurally, SurA contains three domains that are connected by flexible linkers including a core chaperone domain comprised of both the N- and C-terminal regions of the sequence, and two peptidyl prolyl isomerase (PPIase) domains, P1 and P2. Figure 2.1A shows the monomeric crystal structure of SurA in which the P1 domain is bound to the core domain while the P2 domain is extended away from the core-P1 complex (67). In solution, however, alternative conformations of SurA are thought to exist. Indeed, the chaperone activity of SurA has been linked to the conformational dynamics of its domains, as a mutation at the core-P1 interface was shown to increase chaperone activity *in vivo* (40). These data imply an "open" conformation of SurA in which each domain is

structurally isolated from the others; this open conformation has subsequently been confirmed as the active, uOMP-binding conformation of SurA (69, 108).

In this work, we investigated the intrinsic properties of SurA in solution to better understand and model SurA function in the uOMP biogenesis pathway. Unlike the other general chaperones in the uOMP biogenesis pathway that form functional oligomers we find that SurA is monomeric in solution at concentrations equivalent to the reported expression level in E. *coli*. Using small angle neutron scattering to examine the hydrodynamic properties of SurA we observe that the radius of gyration is compatible with an average domain organization in which one PPIase domain is unbound and extended away from the two interacting domains. We further used the experimental scattering curve to assess which structural models of SurA are representative of the solution conformation of SurA. This analysis potentially identifies a novel conformation of the P1-closed conformation in which the P1 domain is docked to the core domain using an interaction surface not observed in crystal structures. Finally, we measured the energetics associated with PPIase-core domain interactions and discovered that both the P1 and P2 domains compete for binding to the core domain. Together, our results quantify the intrinsic conformational ensemble of SurA and provide insight into a potential mechanism for the regulation of its chaperone function.

2.2 Methods

SurA Construct Cloning. The WT SurA plasmid was designed by inserting the *E. coli* gene for SurA with a C-terminal 6-Histidine tag into the pET28b vector between restriction sites Nde I and BamHI. The signal sequence was omitted from the sequence for cytoplasmic expression. Primers for SurAΔP1 and core domain constructs were designed using the Takara Infusion primer design website tool (https://takara.teselagen.com/#/DesignPage). These two constructs were cloned using the In-Fusion HD Cloning Plus CE method (Takara). SurAΔP2 was created through a multi-step approach in which both the first 281 amino acids of SurA were amplified (N and P1 domains) and the region starting with residue 383 until the end of the sequence were amplified. These two fragments were joined and amplified using PCR, followed by insertion into a linearized pet28B vector using Gibson assembly. All primers used to create SurA constructs are listed in Table 2.1.

Stellar cells were transformed with the PCR product by heat shock and plated grown overnight on LB Plates with 50 μg/mL kanamycin at 37°C. Plasmid DNA was extracted from single colonies with the GeneJET Plasmid Miniprep Kit and sequenced to validate the construct sequences. Plasmids were transformed into E. coli HMS174(DE3) via electroporation for protein expression and stored at -80^{o}C as glycerol stocks.

SurA Expression and Purification. 5mL Terrific Broth (TB) cultures containing 50 μg/mL kanamycin were inoculated with cells containing a SurA plasmid and grown overnight at 37°C. These cultures were used to inoculate a 500 mL TB culture with antibiotics at 37°C and were grown until reaching an OD_{600} of 0.6-0.8, followed by addition of IPTG to induce

expression our constructs overnight. Cells containing SurA FL and SurA ΔP2 plasmids were expressed at 37°C, while cells containing plasmids encoding for SurA ΔP1 and the isolated core domain were expressed at room temperature. For all growths, cells were pelleted by centrifugation at 5000 rpm for 15 minutes (Beckman J2-MI, JA-10 rotor) the following morning and stored at -20°C.

Cells were thawed for lysis and then solubilized in 25 mL of Buffer A (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 8.0) containing a EDTA-free protease inhibitor tablet (Pierce). An Emulsiflex homogenizer (Avestin) was used to lyse cells, followed by centrifugation (Beckman J2-MI, JA-10 rotor)for 30 minutes at 5500 rpm to pellet cell debris. Clarified cell lysate was passed through a .45 μM Millex filter prior to being loaded onto to a Ni-NTA Sepharose High Performance bench-top column that was preequilibrated with Buffer A. The column was then washed with 40 mL of Buffer A, followed by elution with 30mL of Buffer B (Buffer A + 300 mM imidazole). A second round of protein purification was performed using FPLC with a Superdex 75 Increase 10/300 GL column (GE) in an elution buffer of 20 mM Tris-HCl, pH 8.0. The purity of elution fractions was assessed by SDS-PAGE and protein concentration was assessed using the theoretical extinction coefficient 29450 M^{-1} cm^{-1}. SurA stocks were stored at -20°C until experiments were performed.

Sedimentation Equilibrium AUC. Sedimentation equilibrium analytical ultracentrifugation (SE) experiments were used to evaluate the oligomerization state of SurA in solution. SurA was diluted into three samples with A_{280} = 0.90, 0.60, and 0.30 at a path length of 1.2 cm (corresponding to concentrations of 25 μM, 17 μM, and 8 μM

respectively) in 20 mM Tris buffer (pH=8.0). Samples were loaded into six-sector centerpieces and allowed to equilibrate at speeds of 20,000 rpm, 24,500 rpm, and 30,000 rpm in a Beckman Optima XL-A analytical ultracentrifuge with absorbance optics. Data were collected at 37 °C with radial scans (wavelength, λ = 280 nm) acquired with 0.001 cm radial steps with 10 replicates. The condition of sedimentation equilibration was confirmed using WinMatchv0.99 and data were subsequently trimmed using WinReedit v.0999.0028 to regions where Beer's law applies (109). Global fitting was completed utilizing WinNonLin v.1.06 (109). For data analysis, we calculated partial specific volume values and buffer densities using Sednterp v.20130813b.(110) The values used for protein partial specific volume ($\bar{v}$) is: $\bar{v}$ = 0.7325 mL g^{-1}. For 20 mM Tris buffer (pH=8.0), the buffer density (ρ) and buffer viscosity (η) are ρ = 0.9988 g mL^{-1} and η = 1.0069 mPa.

SANS of WT SurA. All scattering experiments were collected at the National Institute of Standards and Technology Center for Neutron Research (Gaithersburg, MD) as previously described (111). The scattering data presented here were collected on the NGB 30-m SANS Instrument (NIST). A neutron beam of wavelength λ = 6 Å (wavelength spread, $\Delta\lambda/\lambda$= 0.15) was utilized to collect scattering profiles from all samples described here on a 2D position-sensitive detector (64 cm x 64 cm) with 128 x 128 pixels at resolution of 0.5 cm pixel^{-1}. For data processing, raw counts were normalized to a common monitor count and then corrected for empty cell counts, ambient room background counts, and non-uniform detector response. Data were placed on an absolute scale by normalizing the scattering intensity to the incident beam flux for each individual pixel. Radial averaging was utilized to produce scattering intensity profiles, I(q) versus q;

$q = 4\pi\sin(\theta)/\lambda$, where 2θ is the scattering angle, λ is the neutron wavelength, and q is the magnitude of the scattering vector. Sample-to-detector distances of 5.0 m and 1.5 m were used to cover a range of 0.01 $Å^{-1} < q < 0.4$ $Å^{-1}$.

We prepared the WT SurA protein with a slightly modified protocol for SANS analysis. After expression and purification in that we further purified SurA by gel-filtration in 20 mM Tris and 200 mM NaCl (pH=8.0, GF buffer). SurA (40 μM) was injected onto a Superdex 200 10/300 (GE Healthcare Life Sciences) gel-filtration column in GF buffer with a flow rate of 0.6 mL/min. Fractions containing SurA were pooled and buffer exchanged into GF buffer containing 98 % D_2O via centrifugation in an Amicon filter (Millipore) with a 10-kDa MWCO. We collected the scattering profile of apo-SurA at 1 mg mL^{-1} (20 μM).

For initial analysis of SANS data, we utilized the Guinier approximation to obtain two fit parameters: the macromolecule R_G (Å) and the forward scattering intensity at q = 0 (*i.e.*, I(0) in cm^{-1}). This approximation estimates the intensity in low q regions as follows:

$$I(q) \approx I(0)\exp\left[-\left(\frac{1}{3}\right)R_G^2 q^2\right] \qquad \text{(Equation 2.1)}$$

$$\ln[I(q)] \approx \ln[I(0)] - \frac{1}{3}R_G^2 q^2 \qquad \text{(Equation 2.2)}$$

Therefore, linear regression of ln[(I(q)] vs. q^2 yields information in the slope (*i.e.* R_G^2) and the intercept (*i.e.*, I(0)). I(0) was also calculated using the Contrast Calculator module (112) in the web version of the SASSIE software developed at NIST (113):

$$I(0) = \frac{C\Delta\rho^2\bar{\nu}^2 M}{N_A} \qquad \text{(Equation 2.3)}$$

In the above equation, C indicates the protein concentration in mg mL^{-1}, $\Delta\rho$ is the contrast, $\bar{\nu}$ is the protein partial specific volume in mL g^{-1}, M is the molecular weight in Da, and N_A is Avogadro's number. We compared the fit value of I(0) to the calculated value to ensure that samples contain homogeneous, monomeric species (Table 2.2).

SANS SasCalc Calculations. This module requires several input parameters to be specified including PDB files to indicate atom positions, buffer (H2O:D2O) composition, protein deuteration level and I(0) value. For SasCalc, we built the missing linkers/loops and a C-terminal histidine tag onto 1M5Y using Modeller (114). We created the open conformation by combining the core-P1 domain arrangement from 2PV3 (68) and the core-P2 domain arrangement from 1M5Y. The P2-closed conformation was built by taking the open conformation and collapsing P2 to the binding groove between the core and P1 domains.

Evaluation for Agreement between SANS Data and Structural Models. The SasCalc module in SASSIE was used to calculate SANS scattering profiles ($P(q)_{calc}$) for all structural models (113, 115). SasCalc curves were evaluated for their ability to describe the experimental SANS curve using the reduced χ^2 as recommended by Trewhella and colleagues (116):

$$\chi^2 = \frac{1}{N-1}\sum_{1}^{N}\left[\frac{I(q)_{obs} - I(q)_{calc}}{\sigma_{obs}}\right]^2 \qquad \text{(Equation 2.4)}$$

where N equals the number of data points, $I(q)_{obs}$ and $I(q)_{calc}$ are the experimental and calculated intensity values, respectively, at each point q and σ_{obs} is the error on the experimental measurement at each point. A good fit is defined as $\chi^2 = 1$.

CD Urea Titrations. CD titrations were collected on an Aviv 62A DS spectropolarimeter using a 0.1 cm quartz cuvette. SurA and domain deletion constructs were diluted to 1 μM (SurA, SurAΔP1, SurAΔP2) or 1.5 μM (core) in buffer containing 20 mM Tris (Fisher Scientific), pH 8.0. Equilibrium unfolding titrations were conducted by titrating a solution containing equivalent protein concentration to the analyte and 8 M urea (ThermoFisher), 20 mM Tris, pH 8.0 using a computer controlled titrator (Hamilton) to maintain constant protein concentration. Each urea step was between 0.1 and 0.2 M urea until a final concentration of 7 M urea was reached, with 5 min of equilibration time with stirring between each reading. Signal at 222 nm was averaged at each data point 30 seconds with stirring off. Three repeats were conducted for each construct.

We then used a two-state linear extrapolation model (117) to fit the normalized titration curves with the equation below:

$$Y_{obs} = \frac{(Y_N + a_N[Urea]) + (Y_U + a_U[Urea]) * e^{-\left(\frac{\Delta G^O_{N-U} + m[Urea]}{RT}\right)}}{1 + * e^{-\left(\frac{\Delta G^O_{N-U} + m[Urea]}{RT}\right)}} \qquad \text{(Equation 2.5)}$$

where Y_{obs} is the observed signal, Y_N and Y_U are the intercepts of the native and unfolded baselines respectively, a_N and a_U are the slopes of the folded and unfolded baselines respectively, m_T is the m-value, [Urea] is the concentration of urea, $\Delta G_{N\text{-}U}{}^o$ is the change in Gibbs free energy of folding, R is the gas constant, and T is the temperature (393K).

The m-value was determined by globally fitting all three repeats: m = 1.78 for SurA and SurAΔP2, m = 1.75 for SurAΔP1 and SurA core. Mean values and standard deviations were calculated for the ΔG values of each construct.

2.3 Results

SurA is monomeric in solution. Many chaperones self-associate to form higher order species that bind to their unfolded client proteins (118–120). The other general chaperones in the uOMP biogenesis pathway, Skp and FkpA, both use this oligomeric approach to solubilize uOMPs in the periplasm (60, 111, 121, 122). SurA, on the other hand, is thought to exist in a monomeric conformation in the periplasm. However, SurA crystallizes as a dimer when the P2 domain is deleted (68). It was unclear until now whether full-length SurA can dimerize, and whether a dimeric state is populated at the native concentration of SurA in the periplasm.

To determine the oligomeric state(s) of SurA that exist in solution we measured the apparent molecular weight of WT SurA using analytical ultracentrifugation sedimentation equilibrium (SE). Figure 2.1B shows that the SE profile for WT SurA is well described by a single-ideal species model corresponding to the molecular mass of the SurA monomer (global fit shown in Figure 2.2). Notably, this experimental concentration range, i.e. 8 to 25 μmoles/L (μM), encompasses the reported SurA periplasmic concentration (20 μM) (123, 124). These results therefore suggest that WT SurA is monomeric under physiological protein concentrations.

SurA Exists in an Expanded Conformation in Solution. Given that SurA is monomeric in solution, we investigated the structural conformation of SurA in solution to gain insight into how this may affect its function. Specifically, the relative orientation of the domains of SurA has been recently called into question. In the monomeric crystal structure of SurA reported by Bitto and McKay (PDB ID: 1M5Y) the P2 domain exists in

an extended conformation away from the core and P1 domains, which are interacting (67). We refer to this general domain arrangement as “P1-closed”. In contrast to this domain organization, Radford and coworkers recently used a combination of chemical crosslinking and single molecule FRET (smFRET) to show that the P2 domain may exist in a more collapsed conformation in solution (69).

To address the discrepancy of the overall size and shape of SurA between the two studies, we measured the solution hydrodynamic properties of SurA using small angle neutron scattering (SANS). SANS is a label-free method that reports on the intrinsic size and shape of biomolecules in solution. Guinier analysis of the scattering profile of SurA reveals the radius of gyration (R_G) for WT SurA equals 32.8±0.5 Å (Figure 2.3A, Table 2.2). This finding agrees with R_G of the monomeric 1M5Y SurA crystal structure calculated using HullRad (33 Å)(125) and indicates that the unbound PPIase domain is located distal from the core-PPIase complex on average.

We next used the entire SANS scattering curve to evaluate whether any of the available 22 structural models of SurA represents the average SANS curve in solution. The available models are derived either from crystal structures or molecular dynamics simulations. Although the monomeric crystal structure R_G agrees with the Guinier region, a reduced χ^2 analysis using the model SANS curve calculated from the monomeric crystal structure (1M5Y) using the SasCalc module in SASSIE (113, 115) shows that this crystal conformation is not in fact a good representation of the data as evidenced by the lack of overlay between the predicted and observed scattering curves (Figure 2.3B), by nonrandom residuals (Figure 2.4), and by a reduced χ^2 = 4.567 (Table 2.3) that is

significantly elevated above the target value of $\chi^2 \cong 1$ for a good fit. We observed much poorer fits for "open" (χ^2 = 10.063) and "P2-closed" (χ^2 = 10.985) conformations of SurA monomers. The P2-closed conformation has a domain arrangement in which the P2 is bound to the core domain while the P1 domain is extended away from the core-P2 complex, and its existence is implied by the folding studies described below. Finally, 17 of 19 structural models recently published also show poor fits (Table 2.3) even though these structural models more broadly explore the conformational arrangements of the P1 and P2 domains (69). Figure 2C depicts a succinct summary of how all 22 structural models describe the data: those with predicted R_G values less than the experimental R_G display large reduced χ^2 values and do not fit the data.

Having eliminated most structural models, we found two published models that describe the experimental SANS curve well (reduced $\chi^2 < 2$). Figure 2.3B shows the overlay of the SasCalc curve for the best fitting model, termed P1C1 (**P1**-**c**losed number **1**) for which χ^2 = 1.536. Figure 2.3D shows this structural model oriented similarly to the monomeric crystal structure shown in Figure 2.1C. Although not shown, both the P1 orientation and the SasCalc curve for the second structure, P1C2, (χ^2 = 1.655) are very similar. P1C1 and P1C2 are similar to the monomeric crystal structure in that all three of these can be generally considered as P1-closed conformations because the P2 domain is isolated in solution and the P1 domain is docked onto the core domain. However, the core-P1 interaction in P1C1 is distinct from the monomeric crystal structure: it occurs through a different interface than the crystal structure. In P1C1, the P1 domain has translated and rotated relative to the core domain. Overall, the abilities of these particular

SurA conformations to fit the experimental scattering profile better than any other structural model means that a predominantly P1-closed conformation represents the ensemble average structure of SurA in solution.

The P1 and P2 Domains Independently Compete for Binding to the Core Domain. To gain insight into the conformational ensemble of SurA, we measured the binding energetics of P1 and P2 domains to the core domain. To access these interaction energies, we first measured the thermodynamic stability of multiple SurA domain-deletion constructs using chemical denaturation titrations. Figure 2.5 shows that the denaturation titrations of each construct are well described by a two-state, linear extrapolation model (126). In these titrations, we monitored the urea-dependent change in the circular dichroism (CD) signal at 222 nm, which reports on the foldedness of α-helices. This signal reports primarily on the core domain because it contains the vast majority of the α-helices in SurA (Figure 2.4). The two-state behavior of every SurA construct further confirms that CD is primarily measuring the folding of just the core domain, as more transitions would be expected in the cases where the PPIase domain unfolding was visualized in our experiments.

These data show that each SurA construct has a different stability. Because our titrations report on the core domain in each construct, these results indicate that the presence of the P1 and P2 domains modulate the stability of the core domain through favorable interactions. The intrinsic interaction energies between the core and PPIase domains can be calculated by taking the difference between the stabilities of the isolated core domain from those of either SurAΔP2 or SurAΔP1, e.g. $\Delta G^{o}_{P1,int} = \Delta G^{o}_{F,SurA\Delta P2} -$

$\Delta G^o_{F,SurA\Delta P1\Delta P2}$ and $\Delta G^o_{P2,int} = \Delta G^o_{F,SurA\Delta P1} - \Delta G^o_{F,SurA\Delta P1\Delta P2}$. Using this strategy, we find a thermodynamically favorable core-P2 interaction, $\Delta G^o_{P2,int} = -0.6$ kcal mol^{-1}. This was surprising because an interaction between the core and P2 domain has not been structurally resolved, though recent work suggests that the P2 domain may reside closer to the core domain than in the crystal structure (69). In contrast, the P1 domain interacts slightly more favorably, $\Delta G^o_{P1,int} = -1.47$ kcal mol^{-1}, as expected from its docked conformation.

The ability of the P2 domain to bind the core domain raises the question of whether the PPIase domains can both bind the core domain at the same time or if they compete for a single binding site. To address this question, we constructed a thermodynamic cycle of the four SurA domain-deletion constructs as shown in Figure 2.6. This analysis reveals that both PPIase domains bind the core domain more favorably in the absence of the other PPIase domain. In other words, the presence of the P2 domain affects the apparent affinity of the P1 domain and vice versa. These differences are reflected in the $\Delta G^o_{P1,comp}$ and $\Delta G^o_{P2,comp}$ energy terms, and this result means that the P1 and P2 domains compete with each other for a binding site on the core domain. The nature of the competition can be assessed by comparing the summed free energy changes along the two paths from the WT SurA (top-left corner of the cycle) to the isolated core domain (bottom-right corner). Figure 2.6 demonstrates that these two paths (curved arrows) are energetically equivalent. Therefore, there is no evidence for allosteric communication between the P1

and P2 domains and that the two microscopic binding events are thermodynamically uncoupled and thus independent.

The observation of independent, competitive binding of the P1 and P2 domains to a single site on the core domain necessitates that at least three distinct conformations of SurA exist in solution. Figure 2.7 shows the distribution of these populations at equilibrium calculated from the interaction energies. The dominant conformation (75%) is P1-closed and the second most favorable conformation (18%) is P2-closed. To switch between these two conformations, an “open” conformation where both the P1 and P2 domains are structurally isolated from the core domain must exist in the remaining 7% of the population of the SurA conformational ensemble. These equilibrium populations are in agreement with the SANS data that revealed the P1C1 P1-closed conformation to be most represented of the average ensemble in solution.

2.4 Discussion

SurA is a key member of the periplasmic chaperone network in the OMP biogenesis pathway in gram-negative bacteria. The other chaperones in this network, Skp and FkpA, oligomerize to sequester uOMPs to prevent their aggregation and shield them from interacting with other proteins (111, 121, 127–130). In contrast, we find that SurA does not oligomerize in solution at physiological concentrations. This means that SurA does not adopt a structural organization that encapsulates uOMP clients, consistent with the distinct roles that SurA plays in the OMP biogenesis pathway compared to Skp and FkpA. Unlike these chaperones, evidence suggests SurA delivers uOMPs to the BAM complex to initiate their folding (66). Additionally, recent work has shown that SurA can interact with uOMPs as they are secreted into the periplasm through the translocon (131). These auxiliary functions require a binding mode in which SurA can protect the client uOMPs while allowing for uOMPs to interact with other proteins in the biogenesis pathway.

Indeed, the conformational dynamics of SurA have been proposed to regulate and promote binding of client uOMPs (69). In particular, mutations affecting the core-P1 interaction increase the apparent chaperone activity of SurA in vivo (40), and the transition between the “P1 closed” and “open” conformations has been shown be necessary for uOMP binding. Using equilibrium thermodynamics, we have found that both the P1 and P2 domains compete for binding to the core domain. This competition results in 93% of the SurA molecules existing in a conformation in which one of the PPIase domains of SurA occupies the proposed uOMP binding site. Therefore, the P1 and P2 domains inhibit the exposure of the client-binding site on SurA in the absence of uOMPs.

As SurA is thought to only bind a subset of OMPs (56), we speculate this inhibitory function may serve to prevent SurA from binding non-client uOMPs or other proteins in the periplasm. Interestingly, this autoinhibition does not appear to allosterically regulated, although it should be noted that our work does not exclude cooperative coupling in the presence of a client uOMP. Further structural studies will be required to evaluate the details of whether there is a single binding site or two overlapping binding sites for the P1 and P2 domains on the core domain.

Finally, we investigated the intrinsic hydrodynamic properties of SurA, as the conformation adopted by the unbound PPIase domain (usually P2) may be important for the function of SurA. We find that the R_G of SurA in solution is approximately equal to that of a “P1-closed” domain arrangement in which the P2 domain must also be extended away from the core-P1 complex. Our finding is seemingly in conflict with recent work that concluded that the P2 domain exists in a more collapsed state (69). This contradiction can be resolved by recognizing that the latter conclusions were primarily based on the ability of the P2 domain to chemically crosslink with the core and P1 domains and on smFRET experiments measuring apparent distances between domains. While the presence of crosslinks confirms that the P2 can exist in a collapsed conformation, one drawback is that these data do not report on the relative populations. Moreover, Sosnick and coworkers have shown that the fluorescent dyes attached to intrinsically disordered proteins in smFRET experiments can lead to reduced hydrodynamic sizes as compared label-free methods, such as scattering (132). Because the P2 domain is connected to the core and P1 domains by flexible linkers, it is reasonable to expect the published smFRET

data may overestimate the compaction of the P2 domain in solution. Our SANS-based R_G was obtained in the absence of exogenous modifications of SurA and therefore is a more accurate measurement of the intrinsic size of SurA in solution. Together, the data support a model where the position of the P2 domain relative to the core-P1 complex is highly dynamic with its average position being found in an extended conformation.

Using the full SANS scattering curves, we identified the P1C1 conformation of SurA as best representing the average conformation of SurA in solution. In this structural model the P2 domain is structurally expanded away from the core-P1 complex. The P1 domain in these models is contacting the core domain, but in a location that is different from the canonical monomeric crystal structure. This raises the idea that the PPIase domain binding site on the core domain may be larger than previously indicated from structural studies. The location of the P1 domain in these models warrants further investigation, as it may provide insight into biologically important SurA conformations.

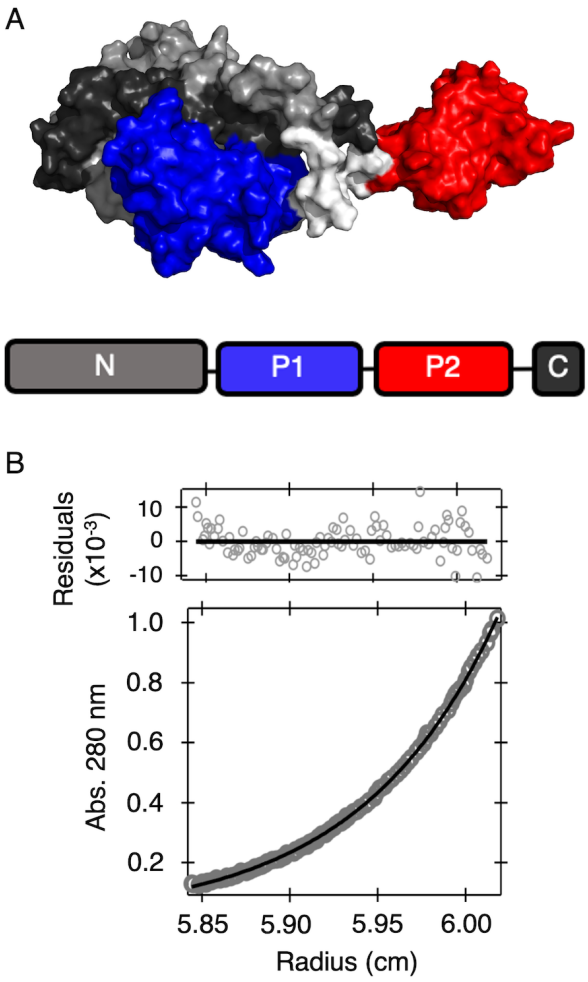
By quantifying the dynamics and energetics that determine the intrinsic conformational ensemble of SurA, we can begin to understand exactly how SurA recognizes uOMPs in the periplasm. In conclusion, we have determined that SurA exists as monomer in solution in a slightly expanded conformation where its PPIase domains compete for binding to the core domain. Paradoxically this dynamic conformational ensemble reduces the population of the chaperone-active, “open” conformation in the absence of client uOMPs. On the other hand, autoinhibition has the advantage of increasing discrimination of SurA for its clients and for preserving the reservoir of apo-SurA in the periplasm.

2.5 Acknowledgements

No author has a conflict of interest. We thank the Center for Molecular Biophysics for providing facilities and resources. We also thank Kathryn Weaver for help cloning the SurAΔP2 construct. Access to NGB30 SANS was provided by the Center for High Resolution Neutron Scattering, a partnership between the National Institute of Standards and Technology and the National Science Foundation under Agreement No. DMR-1508249. We acknowledge the support of the National Institute of Standards and Technology, U.S. Department of Commerce, in providing the neutron research facilities used in this work. This work benefitted from CCP-SAS software developed through a joint EPSRC (EP/K039121/1) and NSF (CHE-1265821) grant. This work was supported by National Science Foundation (NSF) grants MCB1412108 and MCB1931211. D.C.M. and A.M.P. were supported by NIH training grant T32 GM008403. A.M.P. was supported by the NSF grant DGE 1232825. M.J.L. was supported by a Krieger School of Arts & Sciences Aspire Award. The authors thank lab members for helpful discussions. Certain commercial equipment, instruments, or materials (or suppliers, or software, ...) are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

2.6 Figures

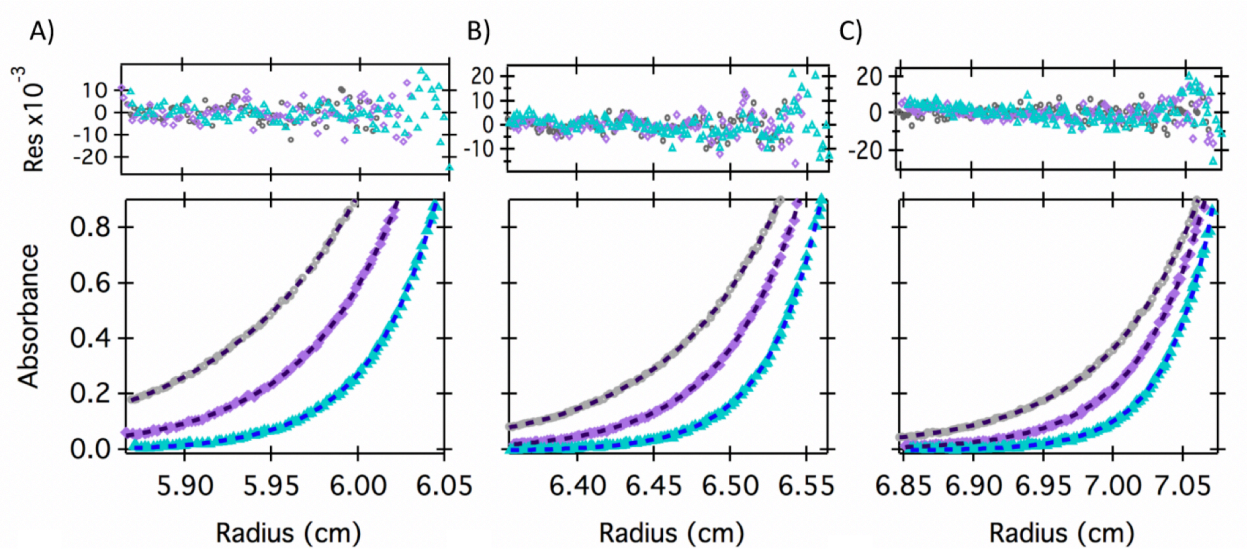
Figure 2.1 - SurA is Monomeric in Solution.



(A) The monomeric crystal structure of SurA (PDB ID: 1M5Y) is shown as a surface representation with its domains colored as depicted in the sequence diagram below. The flexible linkers between the domains of SurA are colored white. In this conformation of SurA, the core (N and C regions) and P1 domains are contacting each other, while the P2 domain is extended away in a structurally isolated conformation. **(B)** Representative SE dataset collected for SurA at a total concentration of 25 μM. These data are well described by a single-ideal species model with a molar mass equal to 43 kDa ± 2 kDa.

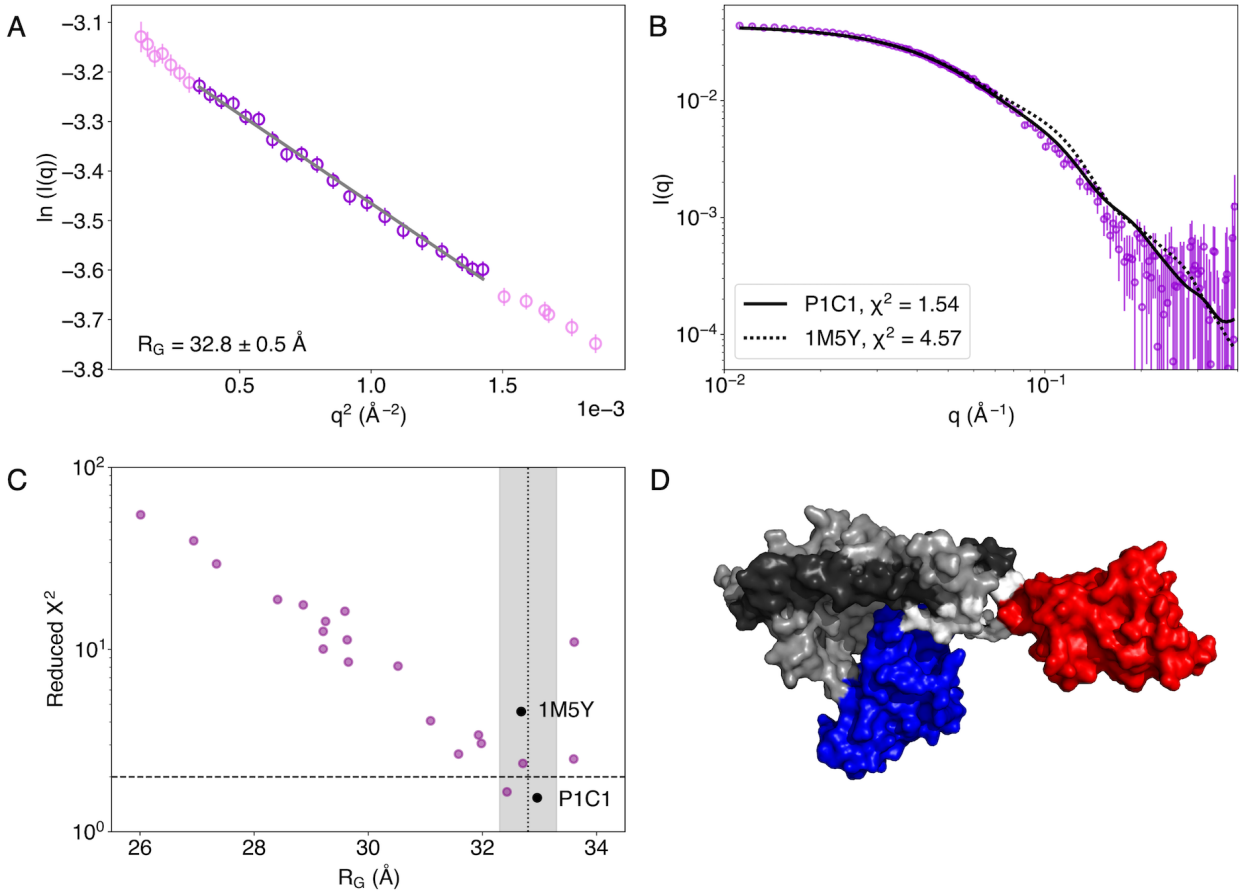
This agrees with the calculated molecular weight of monomeric SurA (35 kDa). These values represent the average weight obtained from fitting three independent experiments and the standard deviation of fitting (Figure 2.2).

Figure 2.2 - Global Fit of WT SurA Sedimentation Equilibrium Data



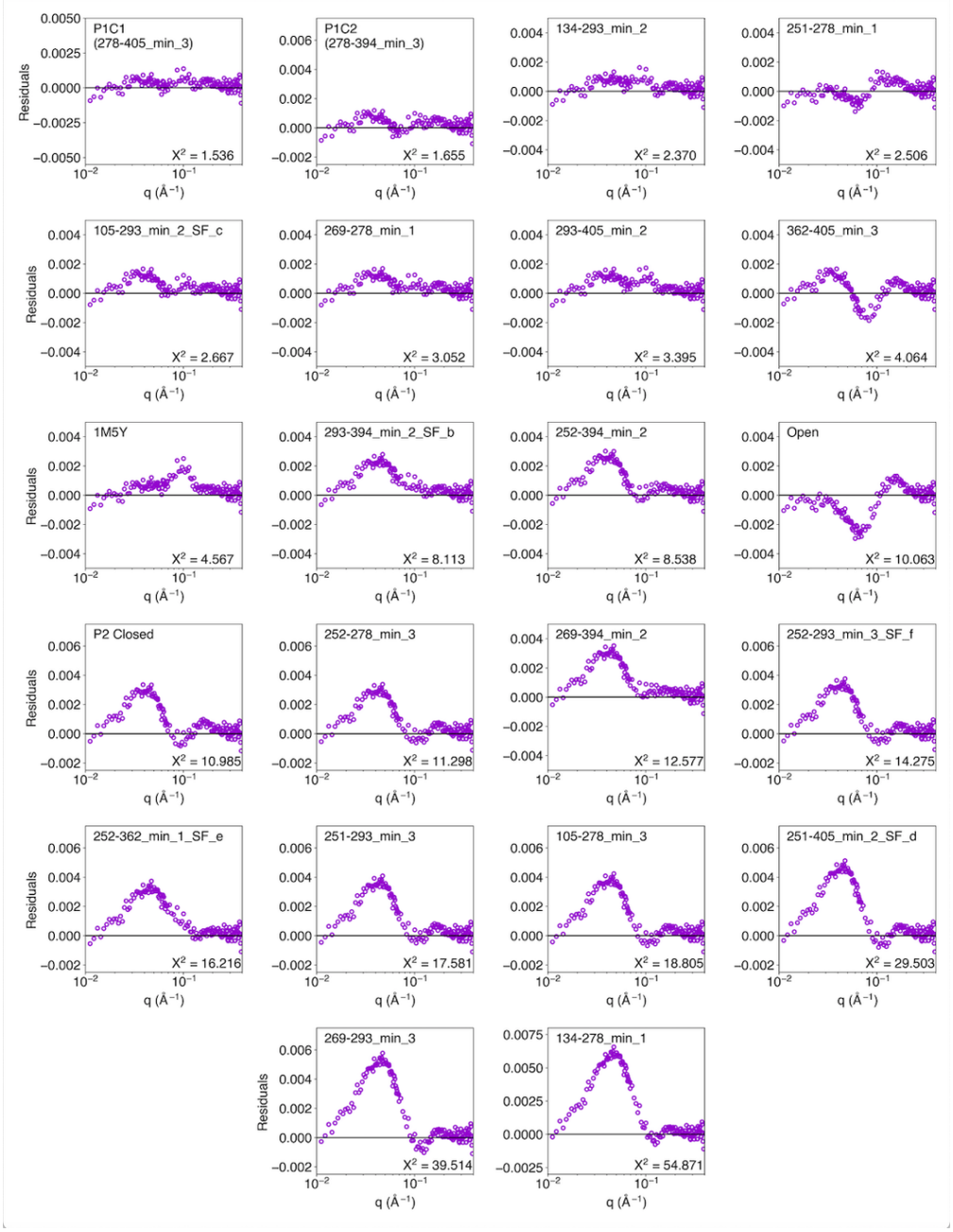
Panels A-C show SE results obtained from data collected in 20 mM Tris (pH=8.0) at three speeds, 20,000 rpm (A), 24,500 rpm (B), and 30,000 rpm (C) and three protein concentrations: of 25 μM (grey), 17 μM (purple), and 8 μM (cyan). Fits to a single-ideal species model are shown as dashed black curves with residuals plotted for each fit. The best fit molar mass equals 43 ± 2 kDa, consistent with the molar mass of monomeric SurA.

Figure 2.3 - Elongated Models of SurA Best Describe SANS Curve.



The Guinier region of the WT SurA SANS dataset in 98% D_2O is shown, with the data shown in darker purple used to conduct the Guinier analysis. The fit for this region is shown with a gray line, with the radius of gyration obtained from the fit shown at the bottom left corner. Additional q*Rg ranges give similar values and are shown in Table 2.2. **(B)** The experimental SANS curve is shown in violet circles with error bars to reflect experimental error. The predicted scattering profiles from P1C1 and 1M5Y are shown as solid and dashed lines, respectively. **(C)** The reduced χ^2 of each available structural model of SurA is plotted against the predicted R_G values calculated using HullRad (125).

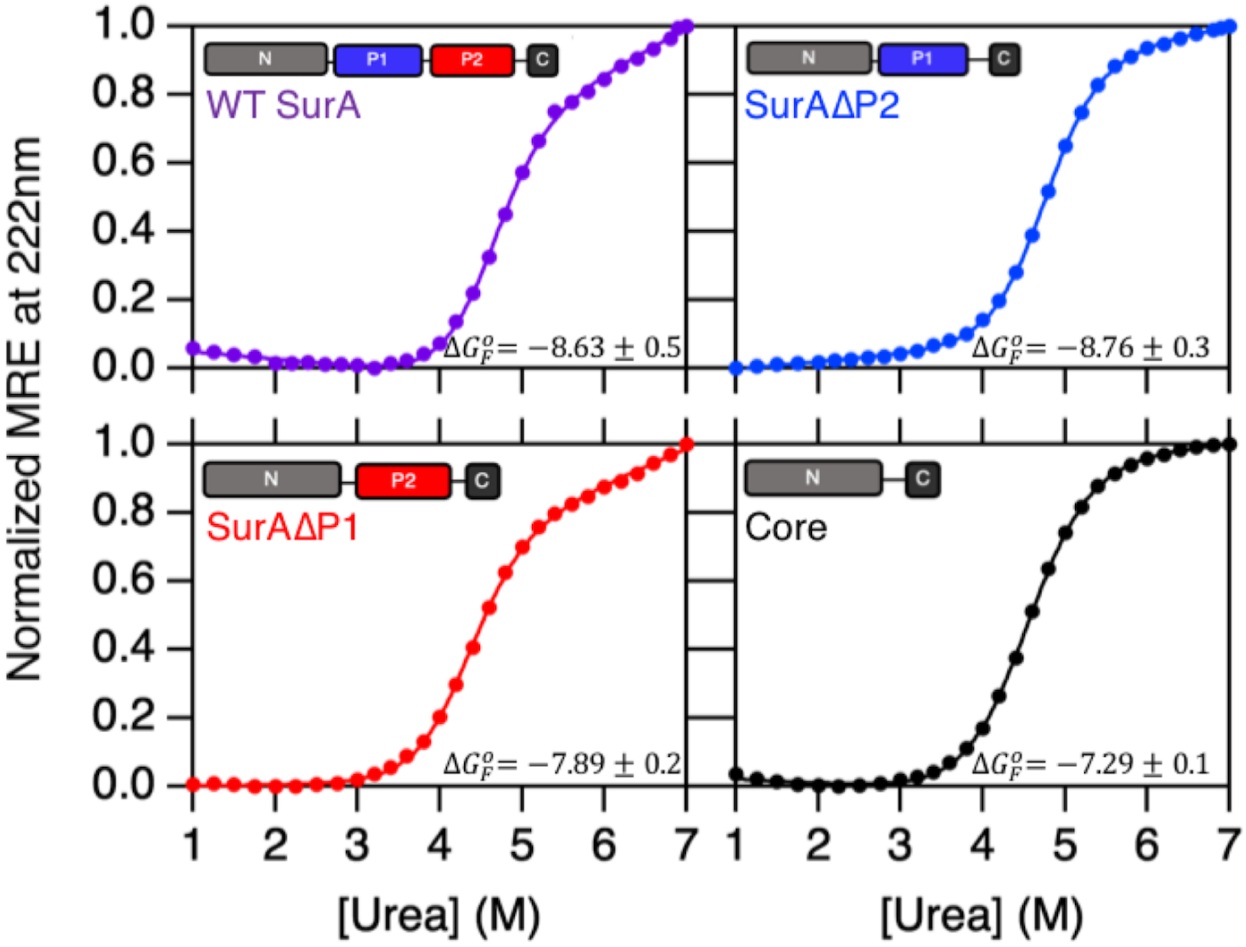
Figure 2.4 - Residuals of SurA model fits to experimental SANS profile.



The residuals for the comparison between the experimental SANS curve for WT SurA and the predicted SANS profile for each available structural model of SurA are shown, ordered from lowest to highest reduced chi-squared. The majority of models show non-

random residuals, indicating that they are poor representations of the average solution structure of SurA. Models sources are given in the legend for Table 2.3.

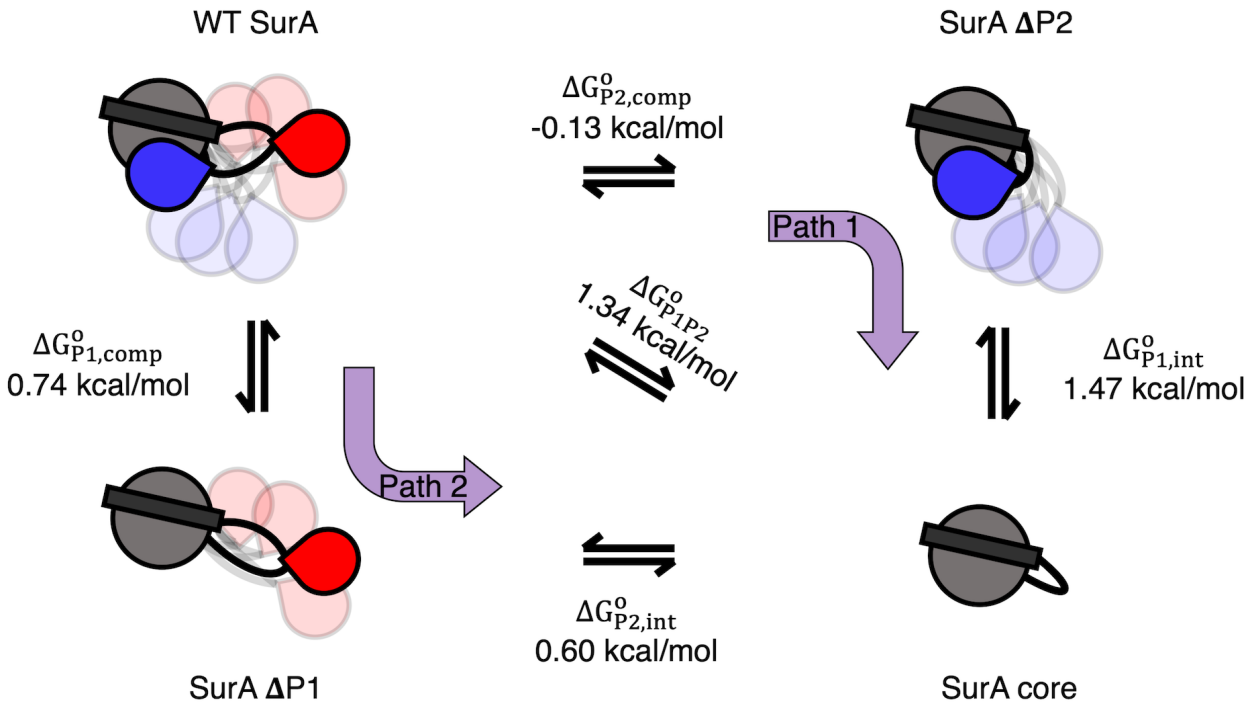
Figure 2.5 - Chemical Denaturation Titrations of SurA Domain-Deletion Constructs.



Circular dichroism signal at 222 nm was monitored as a function the concentration of urea to measure the equilibrium unfolding of SurA constructs of varying compositions. The cartoons in the upper left-hand corner of each plot indicate the domain organization of each construct. Each SurA construct was found to cooperatively unfold and were fit to a two-state, linear extrapolation model. The folding stabilities of each construct are shown in the bottom right-hand corner of each plot and are the average of three independent titrations, with errors representing the standard deviation between the three measured stabilities. The WT SurA and SurAΔP2 titrations were best fit with an *m*-value of unfolding

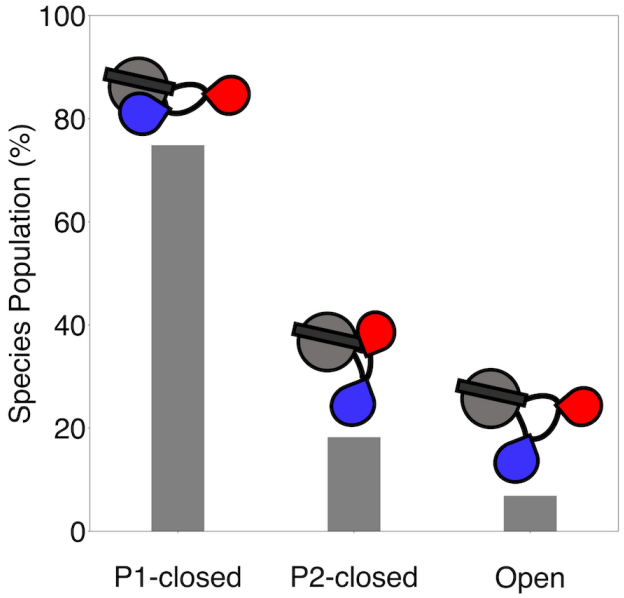
equal to 1.78 and the SurAΔP1 and core domain titrations were best fit with an *m*-value of 1.75, as determined by globally fitting the three titrations for each construct separately.

Figure 2.6 - Thermodynamic cycle analysis reveals competitive interactions between the PPIase domains and the core.



Thermodynamic cycle describing the difference in stabilities between two domain deletion constructs of SurA is shown. The four SurA constructs are shown as cartoons with the most favorable conformation of SurA colored solid and transparent domains to indicate the flexibility of the P1 and P2 domains relative to the core domain. Each side of the cycle is labeled to indicate the corresponding ΔG^{o} in Table 2.4. The two indirect thermodynamic paths from WT SurA to the core domain are indicated with purple arrows, and the direct path is indicated with the diagonal arrows in the middle of the cycle. This analysis reveals that the P1 and P2 domains compete for binding to the core domain in a thermodynamically uncoupled manner. This mechanism necessitates three conformations of SurA.

Figure 2.7 - The relative populations of the three conformations of SurA



Cartoon models of each conformation is shown above the bar. The percent population of each conformation in solution was calculated using $\Delta G^{o}_{P1,int}$ and $\Delta G^{o}_{P2,int}$.

2.7 Tables

Table 2.1 - Primers used to make SurA domain deletion constructs.

Construct	Primer Type	Sequence (5′ – 3′)
SurAΔP1	Forward	ACGACGCCCTGCGCGGCGAAAGCAAAAATATC
	Reverse	CGCGCAGGGCGTCGTTTTGGTTACCCAC
Core	Forward	ACGACGCCACCCGTAATGTCGATAAAACCG
	Reverse	TACGGGTGGCGTCGTTTTGGTTACC
Amplify 1-281	Forward	GCGCCACAGGTAGTCGATAAAGTCGCA
	Reverse	TATCGACATTACGGGTCGAGATATTTTTGCTTTCGCCGCGCAGG
Amplify 383- End	Forward	GCAAAAATATCTCGACCCGTAATGTCGATAAAACCGACGCT
	Reverse	TCAGTGGTGGTGGTGGTGGTGCTCG
SurAΔP2 frag.	Forward	GAAGGAGATATACCCATATGGCGCCACAGGTAGTCGATAAAG
	Reverse	GCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGC
SurAΔP2 Gibson	Forward	ACCACCACCACCACCACTGAGATCCGGCTGCTAACAA
	Reverse	TTATCGACTACCTGTGGCGCCATATGGGTATA

Table 2.2 - SANS Parameters for Guinier Fitting.

I(0) (cm^{-1})	R_G (Å)	$q*R_G$ Range
0.0448 ± 0.0004	32.8 ± 0.5	0.576 - 1.238
0.0448 ± 0.0005	32.8 ± 0.5	0.611 - 1.239
0.0448 ± 0.0005	32.8 ± 0.5	0.646 - 1.239

Summary of the fitting parameters derived from Guinier fitting of SurA data using a range of $q*R_G$ values for analysis. Figure 2.3A shows the Guinier analysis indicated by the row with the gray background. Errors are standard deviations from fitting. For I(0) and R_G values, errors indicate the standard deviations from fitting.

Table 2.3 - Reduced chi-squared and R_G for each structural model of SurA

Structural Model	Reduced χ^2	R_G (Å)
P1C1 (278-405_min_3)[a]	1.536	32.96
P1C2 (278-394_min_3)[a]	1.655	32.43
134-293_min_2[a]	2.370	32.71
251-278_min_1[a]	2.506	33.60
105-293_min_2_SF_c[a]	2.667	31.58
269-278_min_1[a]	3.052	31.98
293-405_min_2[a]	3.395	31.93
362-405_min_3[a]	4.064	31.09
1M5Y	4.567	32.68
293-394_min_2_SF_b[a]	8.113	30.52
252-394_min_2[a]	8.538	29.65
Open	10.063	33.61
P2 Closed	10.985	29.21
252-278_min_3[a]	11.298	29.63
269-394_min_2[a]	12.577	29.21
252-293_min_3_SF_f[a]	14.275	29.25
252-362_min_1_SF_e[a]	16.216	29.59
251-293_min_3[a]	17.581	28.86
105-278_min_3[a]	18.805	28.41
251-405_min_2_SF_d[a]	29.503	27.34
269-293_min_3[a]	39.514	26.94
134-278_min_1[a]	54.871	26.01

The reduced χ^2 and R_G values for each structural model of SurA are shown. Reduced chi-squared values were calculated by comparing the experimental scattering profile of WT SurA to the predicted scattering curve of each model, which was generated by the SasCalc module in SASSIE (113, 115) . Predicted R_G values for each model were calculated using HullRad (125). The two models that best describe the data are labeled P1C1 and P1C2, as they fall into the general category of the P1-closed conformation.

[a] Coordinates from published structural model (69)

Table 2.4 - $\Delta\Delta G^{o}_{interaction}$ Values in Thermodynamic Cycle Analysis

Nomenclature	Initial	Final	$\Delta G^{o}_{interaction}$ (kcal/mol)
$\Delta G^{o}_{P1,comp}$	SurA	SurAΔP1	-0.74 ± 0.53
$\Delta G^{o}_{P2,comp}$	SurA	SurAΔP2	0.13 ± 0.56
$\Delta G^{o}_{P1,int}$	SurAΔP1	Core	-0.60 ± 0.18
$\Delta G^{o}_{P2,int}$	SurAΔP2	Core	-1.47 ± 0.26
ΔG^{o}_{P1P2}	SurA	Core	-1.34 ± 0.50
$\Delta G^{o}_{Path\ 1}$	SurA	Core	-1.34 ± 0.53
$\Delta G^{o}_{Path\ 2}$	SurA	Core	-1.34 ± 0.56

The ΔG^{o} used in the thermodynamic cycle analysis shown in Figure 2.6 are tabulated here. The “initial” and “final” columns indicate the two folding stabilities that were used to calculate the energy change. Errors are propagated from individual construct stability measurements shown in Figure 2.5.

Chapter 3 – SurA is a Cryptically Grooved Chaperone That Expands Unfolded Outer Membrane Proteins

Published as: Marx, D.C., Plummer, A.M., Faustino, A.M., Devlin, T., Roskopf, M.A., Leblanc, M.J., Lessen, H.J., Amann, B.T., Fleming, P.J., Krueger, S., Fried, S.D., Fleming K.G. (2020) *Proceedings of the National Academy of Sciences* 117(45):28026-28035

3.1 Introduction

Proteins must fold into their native three-dimensional structures to perform their functions. For some proteins this folding process is spontaneous and requires no exogenous factors; however, many proteins – particularly, membrane proteins – are predisposed to populate misfolded states or aggregates that are not functional and can be toxic to the cell (133–135). To suppress these pathways, chaperone proteins promote efficient protein folding through interactions with nascent, unfolded proteins (termed clients) (136–140).

One chaperone network of particular importance is the outer membrane protein (OMP) biogenesis network in gram-negative bacteria. Following translocation across the inner membrane, this network solubilizes hydrophobic, unfolded OMPs (uOMPs) in the aqueous periplasm and delivers them to the β-barrel assembly machine (BAM) complex, which catalyzes uOMP folding into the outer membrane (18, 24, 66, 141, 142). OMPs play several critical roles in bacterial physiology such as nutrient uptake, lipid remodeling, and efflux (25). Recently, the OMP biogenesis pathway has been exploited as a target for the development of antibiotics against gram-negative bacteria because drugs that compromise essential OMP maturation need only cross the fairly porous outer membrane and not the tightly regulated inner membrane (143–145).

The uOMP biogenesis chaperone network is comprised of three proteins: SurA, which has been shown to be the most important protein in the pathway, as well as Skp and FkpA (47, 53, 54, 57, 106, 146). SurA handles the majority of the flux of unfolded outer membrane proteins (uOMPs) through the periplasm, and accordingly a Δ*surA* null-

strain induces a pronounced envelope stress response (24, 34, 47–52). Eight OMPs of varying size (8–26 transmembrane β-strands) and sequence composition have been identified as SurA clients because their expression is notably decreased in the Δ*surA* strain (25, 39, 51, 55, 56).

The mechanism by which SurA binds and solubilizes client uOMPs is currently unknown. The lack of ATP in the periplasm implies that the driving forces for SurA's function must derive from interactions with its clients. Unlike the other members of the OMP biogenesis network that oligomerize to form cages around uOMPs, SurA lacks an obvious cavity to shield uOMPs from the aqueous periplasm and does not readily oligomerize in solution (24). Instead, SurA has a modular structure with three distinct domains connected by flexible linkers shown in Figure 3.1A: a primarily alpha helical "core" domain comprised of portions from the N- and C-terminal regions, and two peptidyl prolyl isomerase (PPIase) domains (P1 and P2) (127, 147–149). The orientations of the P1 and P2 domains relative to the core domain have been shown to be dynamic, though how the SurA conformational ensemble contributes to uOMP binding is unclear (40, 57, 69, 150). The core domain of SurA is thought to be responsible for the majority of the binding energy to small, eight-stranded uOMPs and alone can complement the Δ*surA* strain of *E. coli* (55, 57, 152, 58–64, 151).

Structural elucidation of a SurA•uOMP conformation is challenging. As discussed above, SurA has been shown to exist in multiple conformations. Moreover, the unfolded nature of client OMPs poses several additional hurdles that have impeded structural studies using classical techniques: uOMPs lack regular secondary structure, are highly

dynamic, and are prone to aggregation (60, 153–155). We address these challenges by capitalizing on the power of an integrative/hybrid structural biology approach that combines data from crosslinking, mass spectrometry, and neutron scattering to elucidate structural features of the SurA•uOMP ensemble. By incorporating photo-crosslinking unnatural amino acids throughout SurA, we find SurA sites with the highest crosslinking efficiencies to clients line a groove formed between the core and P1 domains. Contrast-matching small angle neutron scattering (SANS) of a crosslinked complex reveals that a canonical uOMP client is greatly expanded when bound to SurA. Mass spectrometry analysis of the photo-crosslinked SurA-uOMP complexes identifies specific segments on client uOMPs that preferentially interact with the SurA groove.

Using our experimental data as restraints, we created a sparse ensemble of 40 configurations of SurA•uOMP complexes. We validated structural features present in this ensemble with additional chemical crosslinking mass spectrometry and SANS experiments that were not included in model generation. We identified three distinct uOMP binding modes and higher order stoichiometries that were sufficient to explain the data. Overall, our findings provide a structural basis for how SurA solubilizes its uOMP clients and provide a template for how future studies might elucidate the structures of highly dynamic chaperone complexes with unfolded proteins.

3.2 Methods

SurA Expression and Purification- We introduced the gene of E. *coli* SurA lacking the signal sequence into the pET28b vector between the NdeI and BamHI restriction sites with a C-terminal 6-Histidine tag. The library of $SurA_{pAF}$ variants were created by incorporating an Amber stop codon (TAG) into 32 individual positions in the gene (QB3 Berkeley Microlab) (Figure 3.2) (Certain commercial equipment, instruments, material, suppliers, or software are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.). Surface exposed sites were chosen using the 1M5Y crystal structure of apo-SurA. The "locked-closed" SurA variant was created by cloning in two point-mutations (P61C/A218C) using InFusion Cloning (Takara). Plasmids were transformed into HMS E. *coli* cells. For *p*AF incorporation, HMS cells also harbored the pDule2 plasmid which encodes the amber-suppressor $tRNA_{Tyr(CUA)}$ from *Methanocaldococcus jannaschii*, and its cognate aminoacyl-tRNA synthetase (gracious gift from the Sondermann lab).

Following an overnight growth, 500 mL Terrific Broth cultures were supplemented with 50 μg mL^{-1} kanamycin (10 μg mL^{-1} streptomycin was also added to pDule2 containing cultures) and induced at $O.D._{600}$ = 0.6-0.8 with IPTG (final concentration 0.1 mmol L^{-1} (mM)). *p*AF was added to a final concentration of 1 mM at the time of induction for $SurA_{pAF}$ variant growths. Cells were harvested after growth overnight (5000 rpm for 30 min, 4°C) in a Beckman J2-MI centrifuge (JA-10 rotor) and pellets were frozen and stored at -20°C.

For purification, cell pellets were solubilized in Buffer A (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 8.0; one Pierce EDTA-free protease inhibitor tablet added, (Thermo Prod # 88266)) and subsequently lysed using an Avestin Emulsiflex homogenizer. Lysate (supernatant) was harvested via ultracentrifugation (5000 rpm for 30 min, 4^{o}C), then filtered (0.22 μm pore size) and purified using a Ni-NTA Sepharose High Performance bench-top column. After loading onto the column, the sample was washed with 20mL of Buffer A containing 6M urea. Samples were refolded on the column and eluted in Buffer A containing 300 mM imidazole. Purified SurA was then dialyzed overnight (10 kDa MWCO Snakeskin dialysis tubing, Thermo Prod #68100) into 20 mM Tris, pH 8.0, concentrated using a 30 kDa MWCO Amicon spin concentrator (Millipore) and stored at -20^{o}C. Stock concentrations were determined with the theoretical extinction coefficient of 29450 M^{-1} cm^{-1} (156).

pAF Crosslinking- 25 μmols L^{-1} (μM) of each $SurA_{pAF}$ variant was mixed with 5 μM uOMP (20 mmoles L^{-1} (mM) Tris, 1 mol L^{-1} (M) urea, pH = 8.0. uOMPs expressed and purified as described previously (157). We chose these conditions because both SurA and $uOmpA_{171}$ are monomeric and soluble at the listed protein and urea concentrations (150, 154). Mixtures were then irradiated with UV light (wavelength, λ= 254 nm) for 5 minutes using a Spectroline MiniMax UV Lamp (Fisher #11-992-662). Aliquots were taken for SDS-PAGE analysis both pre- and post- exposure to UV light. These samples were subjected to electrophoresis using a 4-20% gradient precast gel (Mini-PROTEAN TGX, Bio-Rad) at a constant voltage of 200 V for 35 minutes at room-temperature.

Using ImageJ, densitometry analysis on the loss of density of the $uOmpA_{171}$ band was utilized to quantitate crosslinking efficiency. Crosslinking efficiency values were corrected for the amount of $uOmpA_{171}$ band lost (~20%) when mixed with WT SurA (not containing *p*AF). A representative SDS-PAGE gel for each SurApAF variant and uOmpA171 is shown in Figure 3.3. This same protocol was utilized to assess the crosslinking of $SurA_{pAF}$ variants to uOmpX and uOmpLA.

Perdeuterated OMP Expression, Purification, and Characterization. Deuterium was incorporated into the $uOmpA_{171}$ protein as previously described (111). Briefly, we expressed $uOmpA_{171}$ to inclusion bodies in minimal M9 growth media containing D_2O and deuterated-glucose. Inclusion bodies were isolated and stored in -20 °C. Prior to use, inclusion bodies were thawed and solubilized in 8 M Urea, 20 mM Tris.

To determine the extent of OMP perdeuteration, which is a required parameter for SANS contrast calculations, we utilized one-dimensional proton NMR. We collected 1D ^1H spectra on both protonated and deuterated $uOmpA_{171}$ (50 μM) in 8 M Urea, 20 mM Tris, 10 % D_2O at 35 °C, on a 600 MHz Bruker Avance II spectrometer (Figure 3.4). Water suppression was achieved using a flipback-watergate sequence and a buffer purging pulse was included to minimize the large urea peak. Each spectrum was collected using 128 scans, a recycle delay of 1.5 s, and acquisition time of 150 milliseconds per Free Induction Decay (FID) scan. Data was processed and analyzed using TopSpin 2.1. The spectra were aligned using the amide resonance peaks, and the baseline of the methyl peaks was corrected using a 5^{th} order polynomial. After baseline correction, the methyl peak volumes were integrated using TopSpin, and the integrated intensities of the methyl

peaks in both the protonated and deuterated samples were compared. The loss of intensity in the methyl peaks between the two samples was used to estimate the deuteration level of deut-uOmpA$_{171}$. OMP deuteration was estimated to be 80 %.

SANS on Protonated-SurA/Perdeuterated-uOmpA171 Complex- SurA$_{105,pAF}$ was crosslinked to deuterated-uOmpA$_{171}$ as described above. This complex was further purified via size exclusion chromatography (GE Superdex-200 10/300 GL; flowrate = 0.6 mL/min) in 20 mM Tris, 200 mM NaCl, pH 8.0 (GF buffer), and buffer exchanged into either 0 % or 30 % D_2O for SANS experiments (same buffer components as SEC) (Figure 3.5). We made three attempts to also collect scattering profiles in 80% and 98% D_2O of this complex but the I(0) values from Guinier fitting indicated that these samples contained aggregates. It is known that increased buffer concentrations of D_2O may promote self-association and aggregation of particularly hydrophobic proteins (158).

All scattering experiments were collected at the National Institute of Standards and Technology Center for Neutron Research (Gaithersburg, MD) as previously described.(111)

SANS Data Analysis. For model-independent analysis of SANS datasets, we utilized the Guinier approximation to obtain two fit parameters: the macromolecule R_G (Å) and the forward scattering intensity at q = 0 (i.e., I(0) in cm^{-1}). This approximation estimates the intensity in low q regions as follows:

$$I(q) \approx I(0)\exp\left[-\left(\frac{1}{3}\right)R_G^2 q^2\right] \qquad \text{(Equation 3.1.)}$$

$$\ln[I(q)] \approx \ln[I(0)] - \frac{1}{3}R_G^2 q^2 \qquad \text{(Equation 3.2.)}$$

Therefore, linear regression (i.e., Figures 3.6 and 3.7 of ln[(I(q)] vs. q^2 yields information in the slope (i.e. R_G^2) and the intercept (i.e., I(0)). I(0) was also calculated using the Contrast Calculator module (112) in the web version of the SASSIE software developed at NIST (113). Our complexes contain components with different deuteration levels and contrast, which is take into account where:

$$I(0) = \frac{CM}{N_A}(\sum_i f_i \Delta\rho_i \bar{v}_i)^2 \qquad \text{(Equation 3.3.)}$$

where C indicates the protein concentration in g mL^{-1}, M is molecular weight, N_A is Avogadro's number, $\Delta\rho$ is the component contrast, and $\bar{v}$ is the component partial specific volume mL g^{-1}, is M_i/M and the summation is over the two components of the complex. This sum includes three terms: one for protonated SurA, one for perdeuterated $uOmpA_{171}$, and a cross-term that originates from inter-protein interactions. At 30 % D_2O using this rigorous calculation method, SurA, $uOmpA_{171}$, and the cross-term contribute to the total I(0): 9 %, 48 %, and 43 % respectively, the main contribution from the cross-term coming from the $\Delta\rho$ value of $uOmpA_{171}$.

To better understand the contribution to the total scattering of SurA and $uOmpA_{171}$ in the absence of the cross-term, we approximated the contribution of these components to the total scattering using the following simplified formula:

$$I(0) = \frac{CM}{N_A}\sum_i (f_i \Delta\rho_i \bar{v}_i)^2 \qquad \text{(Equation 3.4.)}$$

At 30 % D_2O using this simplified calculation method, SurA and $uOmpA_{171}$ contribute 16 % and 84 % to the total I(0), respectively.

For each Guinier fit, we compared the fit value of I(0) to the calculated value for each experiment (Table 3.1).

In addition to Guinier fitting, we obtained R_G and I(0) values from distance distribution functions, P(r) vs. r (Table 3.2). These fits were completed using autoGNOM(159) and a range of 0.011 A^{-1} to approximately 0.2 A^{-1}. Fit values reported in Table 3.2 were obtained using the specified D_{max} values from autoGNOM. The P(r) vs. r curve is shown in Figure 3.6. Exploration of a range of D_{max} values near the specified values did not result in a change in the shape of P(r) versus r, except for small changes in the region near D_{max}, as shown in Table 3.2.

Evaluation for Agreement between SANS Data and Form Factors – We evaluated agreement between the SANS scattering profile collected in the 0% D2O condition and the scattering form factors from models using the reduced according to Equation 3.4 (116):

$$\chi^2 = \frac{1}{N-M}\sum_1^N \left[\frac{I(q)_{obs-I(q)_{calc}}}{\sigma_{obs}}\right]^2 \qquad \text{(Equation 3.5.)}$$

where equals the number of data points, M is the number of calculated scattering curves (form factors) used for comparison, and are the experimental and calculated intensity values, respectively, at each point and is the error on the experimental measurement at each point. A good fit is defined as $\chi^2 < 1.05$.

XL-MS of Photo-crosslinked SurApAF-uOMP. Acetonitrile (ACN), Optima formic acid (FA), trifluoroacetic acid (TFA), Tris, and urea were obtained from Fisher Scientific (Hanover Park, IL, USA). LiChrosolv LC-MS grade water was obtained from EMD Millipore Corporation (Darmstadt, Germany). Pierce Trypsin protease (catalog number 90305) and Glu-C (catalog number 90054) protease were also obtained from Thermo Fisher.

Crosslinked samples comprising of 25 μM $SurA_{pAF}$ (with *p*AF at sites: 59, 94, 105, 120, 233, 245, 260, or 424), and 5 μM $OmpA_{171}$ or 5 μM OmpX were reconstituted in 20 mM Tris pH 8.0, 1 M urea, and crosslinked as described in the previous section (typically on a 50 μL scale, ca. 50 μg total scale of protein). Following crosslinking, solid urea was added to a final concentration of 2 M. Trypsin (1 μg/μL stock concentration) was then added (typically 1 μL) to the samples at a 1:50 enzyme/substrate ratio. The samples were digested overnight at 25˚ C, 700 rpm on a thermomixer.

For each protein complex, we analyzed peptides from a standard single-trypsin digest as well as from a double digest with trypsin then Glu-C in serial, and this was conducted in technical duplicate to generate four separate injections for each sample analyzed. For the latter sample, we added 20 mM Tris pH 8.0 to dilute the urea concentration to 0.8 M, whereupon Glu-C (1 μg/μL stock concentration) was added (typically 1 μL) to a 1:50 enzyme/substrate ratio. These samples were then digested overnight at 30˚C, 700 rpm on a thermomixer.

Both singly-digested and double-digested peptides samples were acidified by addition of small volumes of TFA (~1 μL) to a final concentration of 1% (vol/vol). Samples were then diluted with 0.5% TFA to a final volume of 1 mL to facilitate loading into the cartridges.

Solid phase extraction was carried out using Sep-Pak C18 vacuum cartridges (Waters, Milford, MA, USA) according to the following protocol: Cartridges were first conditioned (1 mL 80% ACN, 0.5% TFA) and equilibrated (4x 1 mL 0.5% TFA), before loading the sample slowly under a diminished vacuum (ca. 1 mL/min). The columns were

then washed (4x 1 mL 0.5% TFA), and peptides eluted by addition of 1 mL elution buffer (80% ACN, 0.5% TFA). During elution, vacuum cartridges were suspended above 15 mL conical tubes, placed in a swing-bucket rotor (Eppendorf 5910R), and spun for 2 min at 350 *g*. Eluted peptides were transferred from Falcon tubes back into microfuge tubes and dried using a vacuum centrifuge (Eppendorf Vacufuge). Dried peptides were stored at -80˚C until analysis. For analysis, samples were vigorously resuspended in 0.1% FA in Optimal water to a final concentration of 1 mg/mL.

Chromatographic separation of digests was carried out on a Thermo UltiMate3000 UHPLC system with an Acclaim Pepmap RSLC, C18, 75 µm x 25 cm, 2µm, 100 Å column. Approximately 2 µg of protein was injected onto the column. The column temperature was maintained at 40ºC, and the flow rate was set to 0.300 µL/min for the duration of the run. Solvent A consisted of 0.1% FA in 2% ACN, 98% water, and solvent B consisted of 0.1% FA in ACN. After accumulation of peptides onto the trap column (Acclaim PepMap 100, C18, 75 µm x 2 cm, 3µm, 100 Å column) for 10 min (during which the column was held at 2% solvent B), peptides were resolved by switching the trap column to be in-line with the separating column, and applying a 100 min linear gradient from 2% B to 35% B. Subsequently, the gradient was increased from 35% B to 40% B over 25 minutes, and then increased again from 40% B to 90% B over 5 minutes. The column was then cleaned with a saw-tooth gradient to purge residual peptides between runs in a sequence.

A Thermo Q-Exactive HF-X Orbitrap mass spectrometer was used to analyze the eluting peptides. A full MS scan in positive ion mode was followed by ten data-dependent MS scans. The full MS scan was collected using a resolution of 120,000 (@ m/z 200), an

AGC target of 3E6, a maximum injection time of 100 ms, and a scan range from 350 to 1500 m/z. The data-dependent scans were collected with a resolution of 15,000 (@ m/z 200), an AGC target of 2E5, a minimum AGC target of 8E3, a maximum injection time of 250 ms, and an isolation window of 2.0 m/z units. To dissociate precursors prior to their re-analysis by MS2, peptides were subjected to a stepped HCD with 22%, 25%, and 28% normalized collision energies. Fragments with charges of 1, 2, and >8 were excluded from analysis, and a dynamic exclusion window of 60.0 s was used for the data-dependent scans.

XL-MS of Photo-crosslinked SurApAF Data Analysis. MS data were centroided and converted to the mzML file format using the msConvert application in the ProteoWizard Toolkit (160), and then analyzed for crosslinks using MeroX Version 2.0 (161). *p*AF was added to the amino acid list with a mass of 188.06981084 Da ($C_9H_8N_4O$). The photo-crosslink was added to the crosslink tab with composition of $-N_2$ (-28.006148 Da), a maximum Cα-Cα distance of 30 Å, specificity site 1 as *p*AF, and specificity site 2 as any amino acid. For tryptic digests, protease sites were allowed after arginine and lysine residues, with lysine blocked by proline as a cleavage site. For double digests, protease sites were allowed at arginine, aspartic acid, lysine, and glutamic acid, with lysine blocked by proline as a cleavage site. For both tryptic digests, a maximum of three missed cleavages was allowed (four was allowd for double digests). For modifications, a maximum of two oxidations of methionine was allowed. Searches were conducted using a FASTA file that consisted only of the uOMP in consideration and SurA. Otherwise, MeroX default parameters (for scoring and FDR calculation) were used.

Upon reviewing the output, a MeroX score of 50 was selected as the acceptance cutoff for crosslinked peptide-spectrum matches (PSMs). This corresponds to a FDR cutoff of <0.01; and in some cases to a far lower cut-off (Table 3.3). In numerous situations, a crosslink site to *p*AF could not be pinpointed down to a specific residue within a given peptide-spectrum match because of insufficient fragment ion data. In these situations, if several PSMs were available in which one provided more specific identification of a crosslink site and others included that site as part of larger nonresolvable region, we then merged the data to take advantage of the greater specificity when it was available. In Supplementary Data 1 and 2, the full list of crosslink sites associated with all PSMs are provided. Lines colored purple represent the most specific crosslink site assignable, and lines colored blue have lower resolution of the crosslink site but are consistent with a crosslink at a more specific site. The crosslink sites are compiled across all the $SurA_{pAF}$ variants for each uOMP in a tab labeled 'compiled_ Omp_SurApafs.'

XL-MS of DSBU-crosslinked SurA-uOMP. A 50 μL solution comprised of 20 μM WT SurA combined with 20 μM of either $uOmpA_{171}$, uOmpX, or $uOmpA_{171}$ (P61C/A218C), was prepared in 20 mM NaP_i pH 8.0, 1 M urea. Crosslinking was carried out by adding disuccinimidyl dibutyric urea (DSBU, ThermoFisher) from a 100 mM stock in DMSO to a final concentration of 1 mM. The sample was then mixed, incubated at room temperature with agitation for 30 min, and then quenched by addition of 1 M Tris pH 8.0 stock to a final Tris concentration of 100 mM. Following crosslinking, solid urea was added to a final concentration of 2 M. Trypsin (1 μg/μL stock, Pierce) was added to the sample (ca. 1–2

μL) to a 1:50 enzyme:substrate ratio. The samples were digested overnight at 25 ˚C, 700 rpm on a thermomixer. For samples prepared with a serial trypsin–GluC digest, the trypsinolysis reactions were diluted with 20 mL NaP_i pH 8.0 to lower the final urea concentration to 0.8 M. Then 1–2 μL of GluC (1 μg/μL, Pierce) was added to a 1:50 enzyme:substrate ratio, and the samples were digested again overnight at 30˚C, 700 rpm on a thermomixer. Both single and double digests were then acidified with TFA to a final concentration of 1% (vol/vol), diluted with 0.5% TFA to a final volume of 1 mL, and then desalted by solid-phase C18 extraction columns, as described previously. Preparation of the sample and analysis by nanoLC-MS/MS was conducted identically to the samples generated by photo-crosslinking, as described previously.

XL-MS of DSBU-crosslinked SurA Data Analysis. MS data were centroided and converted to the mzML file format using the msConvert application in the ProteoWizard Toolkit (160), and then analyzed for crosslinks using MeroX Version 2.0 using the software's standard settings for DSBU (161). Notably, MeroX uses a slightly expanded set of crosslink site specificities for DSBU: site 1 is restricted to be lysines and N-termini, whereas site 2 has an expanded specificity to also include serine, threonine, and tyrosine. Identified crosslinked peptides were filtered to an FDR of 1% and used if they had a MeroX score greater than 50 (which in most cases corresponded to an FDR well below 1%). Peptide spectrum matches (PSMs) were pooled between four separate injections (two replicates with trypsin only, two replicates with trypsin/Glu-C serial digest) to assemble a list of PSMs, given in Supplementary Data 3–5 under tabs labeled 'total.' We then merged together these datasets and removed redundancies to create condensed

lists of crosslinks, also provided in Supplementary Data 3–5 under tabs labeled 'combined.' In several cases, the crosslink site could not be uniquely pinpointed, as occurs when numerous nucleophilic residues occur close to one another within a given peptide. These uncertainties in the position of the crosslink site are shown explicitly in the Supplementary Data tables.

To analyze these crosslinks further in light of the structural models of SurA•uOmpA$_{171}$, solvent-accessible Ca-Ca surface-distances (SASDs) between 85 pairs of DSBU crosslink sites between SurA and uOmpA$_{171}$ were calculated using JWalk (up to a maximum SASD of 85 Å, grid size 1 Å). This expanded list of residue pairs was created by calculating all possible crosslinks that could be associated with a PSM with ambiguous linkage sites. For instance, if the PSM could determine the crosslink site on SurA to be position 50, but could not confidently determine if the crosslink site on uOmpA$_{171}$ was position 44 or 49 (corresponding to XL ID 1 in Supplementary Data 3), then we calculated the SASD both between SurA50–uOmpA44 and SurA50–uOmpA49. For assemblies with higher stoichiometry (i.e., (SurA)$_n$•uOmpA$_{171}$ with n = {2,3,4}), we calculated 170, 255, or 340 different SASDs because the identity of the SurA crosslink site could have come from any copy of SurA. To use the previous example of XL ID 1, the SASDs between SurA$_A$50–uOmpA44, SurA$_A$50–uOmpA49, SurA$_B$50–uOmpA44, and SurA$_B$50–uOmpA49 would all be calculated.

Next, for each structural model, a 'short-list' of the most likely crosslink sites was constructed by taking whichever of the possible linkages associated with a given XL ID admitted the smallest SASD, and discarding the others. To use the previous example,

whichever of the four different SASDs was the lowest (and its associated crosslink sites) was used for XL ID 1. This procedure thereby provides for each structural model a set of 46 non-redundant SASDs. The SASDs were converted into harmonic scores using the following rule: all SASDs < 25 Å were awarded a perfect score of 100. Otherwise, SASDs were converted to scores using the function: score = $100 - 0.09 \times (\text{SASD} - 25)^2$. Any score that was negative was converted to 0.0001. Any XL ID that did not have any SASD calculated by Jwalk was inferred to be greater than 85 Å, and was given a score of 0.0001. This scoring algorithm awards high scores to all crosslinks with SASDs ≤ 35 Å, moderate scores to crosslinks with 35 ≤ SASD / Å ≤ 45 Å, and low scores (less than 50) to all crosslinks with SASDs greater than 48 Å. All SASDs and scores can be found in Supplementary Data 6.

A matrix of scores was constructed for all 46 crosslinks in the context of 23 various SurA•uOmpA$_{171}$ structural models. Using the spectral biclustering algorithm with the 'log' method as implemented in scikit-learn with 20 biclusters (4 for structural models, 5 for crosslinks), the rows and columns of the score matrix were permuted to generate an organization that revealed distinct clusters, as shown in Figure 3.8 and discussed in the main text.

Making structural models of SurA - Models of SurA were constructed based on crystal structures 1M5Y and 2PV3 (Table 3.4). In the 1M5Y structure, the core and P1 domains are close together but the P2 domain is extended. In the 2PV3 structure, the P1 domain is moved away from the core domain and rotated relative to 1M5Y but the P2 domain is missing. Residue segments 20-34, 165-171, 387-394, and 428-430 were built into 1M5Y

and six histidine residues were added to the C-terminus using Modeller to create the “P1 closed” form of SurA (162). For the “P2 closed” form of SurA, PyMOL was used to build the P2 domain from the P1 closed form into 2PV3 (163). The P2 domain was then moved into position against the groove formed by the core and P1 domains using NAMD as described below. The “open” SurA model has the core-P1 relative orientation from 2PV3 and the core-P2 relative orientation from 1M5Y. Domains were oriented in PyMOL and linkage conformations were normalized using NAMD as described below. The “Collapsed” SurA model has the core-P1 relative orientation from 1M5Y and the P2 domain was moved into position against the core using NAMD as described below.

The P2 closed, Open, and Collapsed SurA models, initially constructed using PyMOL, were further manipulated to position the domains, remove Van der Waals clash and relax unstructured linkage segments using NAMD with generalized Born implicit solvent electrostatics in the CHARMM22 force field (164). Domains were positioned with targeted distance restraints as implemented in the collective variable module in NAMD (165). Typically, a harmonic potential was placed on the distance between the centers of mass of two groups of CA atoms with a force constant of 1.0 and the force was applied for 50,000 to 150,000 steps.

Making structural models of the SurA·uOmpA171 complex. Four extended uOmpA$_{171}$ segments (residues 2-21, 54-73, 84-104, 112-132) that contain the six SurA-binding segments were independently submitted, along with the open SurA structural model to the protein-protein docking web server HADDOCK (166). These sequence segments were chosen to include those residues that were found to repeatedly crosslink to the high

efficiency $SurA_{pAF}$ variants. Active and passive residues for HADDOCK were chosen from SurA groove (Figure 3.9).

These docked oligopeptides were inspected using molecular graphics to obtain target distances for docking the full length, unexpanded $uOmpA_{171}$ models (s = 1.65) to the open form of SurA. Docking was accomplished in NAMD using the target distances from HADDOCK peptide docking as distance restraints in the collective variables module of NAMD as described above.

The uOmpA-open SurA models were further manipulated to increase the uOmpA D_{Max} to the target of 150 Å that was determined by P(r) analysis of the SANS data. These expansions were accomplished using distance restraints and the collective variable module in NAMD. Short segments of each bound uOmpA that were furthest apart were identified and the two groups of respective CA atoms were forced to a distance of ~150Å with a harmonic potential as described above. A second open SurA model was then docked to exposed, known binding segments (Figure 3.10) of the extended uOmpA. In three cases, a third open SurA was docked to remaining exposed known binding segments. One extended polypeptide of uOmpA was generated with a D_{Max} ~250 Å and four open SurA models were docked to the four main segments on OmpA that displayed high efficiency cross-linking.

In all, twenty-three models containing one docked SurA, thirteen models containing two docked SurA, three models containing three SurA, and one model containing four SurA were built. Physical dimensions of these models are listed in Table 3.5. Values for R_G and D_{Max} were calculated using HullRad(125). All models contained CHARMM

hydrogens and were used to calculate predicted SANS profiles using the SasCalc server (113).

Comparison of Structural Models Experimental 0% D2O SANS Profiles. The SasCalc module in SASSIE was used to calculate SANS form factor profiles $P(q)_{calc}$ for all models (113, 115). Normalized form factor curves were obtained by dividing $P(q)_{calc}$ by the I(0) for a given data set. SasCalc form factors were evaluated for their ability to describe the corresponding, experimental SANS curves using the equation for the reduced χ^2 (equation 3.4) as recommended by Trewhella and colleagues (116). As expected for an ensemble, a good fit was not obtained by a single structural model. Therefore, fractionally weighted, linear sums of SasCalc form factor curves were calculated. These included the appropriate weighting for different scattering contrast values between SurA and uOmpA$_{171}$. The equations for basis set addition for cases where the components in solution do not possess uniform scattering contrast are described next. For a linear addition of two basis form factors, the composite scattering curve predicted for a structural model, $I(q)_{calc}$, is a function of the contrast and weight contributions of each component as follows:

$$I(q)_{calc} = \left[W_i P(q)_{i,calc} + W_j P(q)_{j,calc}\right] \quad \text{(Equation 3.6.)}$$

where the W_i terms correspond to the fractional contrast weighting terms simulated from a combination of weight fractions (F_i). The weighting terms were obtained from contrast factors calculated for each component as follows:

$$W_i = \frac{C_i}{C_i + C_j} \quad \text{(Equation 3.7.)}$$

where $W_i + W_j = 1$, and the C variable in these equations for each component equals:

$$C_i = \frac{c_{Tot} F_i m_i \left(\sum_a^N f_a \Delta\rho_a \bar{v}_a\right)^2}{N_A}$$ (Equation 3.8.)

where c_{Tot}, N, m, $\Delta\rho$, and $\bar{v}$ are the weight concentration, molar mass (g mol^{-1}), contrast values (Table 3.6), and partial specific volume (ml g^{-1}) for each species, respectively f_a in a complex is the mole fraction of the a^{th} component within that complex; and $\sum_1^A f_a = 1$ where A is the number of components with any given complex. For normalized data the c_{Tot} term can be dropped because its value has no effect on the W values. For example, the scattering equation for a pair of component curves containing a one-to-one complex and free SurA in cases of nonuniform scattering equals:

$$C_{SurA} = \frac{F_i m_{SurA} (\Delta\rho_{SurA} \bar{v}_{SurA})^2}{N_A}$$ (Equation 3.9.)

and

$$C_{Complex} = \frac{F_j m_{Complex} \left(f_{SurA_{105}} \Delta\rho_{SurA_{105}} \bar{v}_{SurA_{105}} + f_{OmpA} \Delta\rho_{OmpA} \bar{v}_{OmpA}\right)^2}{N_A}$$ (Equation 3.10.)

Although the $SurA_{pAF105}$-$uOmpA_{171}$ crosslinked sample formed a distinct 1:1 band as assessed by SDS-PAGE, the excess free SurA was not completely separated from this sample by subsequent gel filtration. Repeat experiments of mock sample preparation indicate that this non-covalently bound SurA is present at mole ratios of 0.3 to 1 and must therefore be taken into account in the SANS profile analysis. This technical issue turned out to be fortuitous because it allowed the population of low levels SurA-uOmpA171 to bind additional SurA protomers, thus populating complexes with multiple SurA molecules bound.

To simulate a wide range of linear combinations, we iterated through weight fractions at 1% increments and compared these to the 0% SANS curves using the

reduced χ^2 (equation 3.5 above) with M increased to 2 to account for the additional loss of degrees of freedom. No pairwise combination of form factors resulted in a good agreement with the data as evidenced by acceptable χ^2 and appropriate mole ratios of non-crosslinked to crosslinked SurA.

We therefore extended these equations to three terms for triplets by the addition of a third term in equations, e.g.

$$I(q)_{calc} = \left[W_i P(q)_{i,calc} + W_j P(q)_{j,calc} + W_k P(q)_{k,calc}\right] \quad \text{(Equation 3.11.)}$$

and

$$W_i = \frac{C_i}{C_i + C_j + C_k} \quad \text{(Equation 3.12.)}$$

and

$$W_i + W_j + W_k = 1 \quad \text{(Equation 3.13.)}$$

and the terms are as described above and evaluating using the reduced χ^2 equation above with M increased to 3 to account for the additional loss of degrees of freedom as compared to the paired case. Each weight fraction within a simulated triplet was incremented by 1% to achieve an exhaustive search of model combinations.

Data Availability. All raw data and detailed protocols, including gel images, SANS profiles, and SurA–uOmpA171 models, are available at https://github.com/KarenGFleming/SurAuOmpA. All raw and analyzed mass spectra are available on ProteomeXchange under identifiers PXD021870 and PXD021872. Supplementary Dataset Files 1-6 are available from PNAS: https://www.pnas.org/content/117/45/28026.

3.3 Results

SurA crosslinks preferentially with client uOMPs. We identified regions of SurA involved in binding uOMPs using short-distance crosslinkers across the surface of SurA. To accomplish this, we incorporated the unnatural amino acid, *para*-azido-Phenylalanine (*p*AF), at 32 non-conserved, surface-exposed positions on SurA (Figures 3.1B and 3.2) using amber suppression (167). *p*AF is a 'zero-length' crosslinker, because it forms highly reactive intermediates that crosslink rapidly and non-specifically to any residue within 3–4 Å (though one crosslinking mechanism has been shown to slightly favor reactions with aromatic amino acids) (168, 169). Previously applied in biochemical assays, we report here to the best of our knowledge the first application of *p*AF crosslinking for structural studies (170, 171).

Each SurA variant with a single amino acid substituted for *p*AF (denoted $SurA_{pAF}$) was mixed with one of three uOMPs: clients $uOmpA_{171}$ (the transmembrane domain of uOmpA, see SI Methods), uOmpX, or a non-client uOmpLA as a negative control (56, 118). Samples were exposed to UV light for 5 minutes and crosslinking efficiency was measured by quantitative SDS-PAGE (Figure 3.3 and Table 3.7) (172). Crosslinking efficiencies to the two client uOMPs vary dramatically with position on SurA, with the highest efficiency sites all residing on the core and P1 domains (Figures 3.1B,C). The finding that high-efficiency crosslinking sites for client uOMPs map to the SurA P1 and core domains is consistent with previous experiments that found that removal of the P2 domain did not affect the binding affinity of $uOmpA_{171}$ to SurA (57). In contrast, the non-client uOmpLA showed low crosslinking efficiencies; indeed, only half of the $SurA_{pAF}$

variants could form crosslinks with uOmpLA at all (Figure 3.10 and Table 3.7). In addition to the high-efficiency crosslinking sites on cognate client uOMPs, we observed lower levels of crosslinking to client uOMPs at other *p*AF sites across the surface of SurA. The observed differences in crosslinking efficiencies between client and non-client uOMPs indicate that SurA is inherently able to distinguish client uOMP sequences in solution without the aid of other chaperones.

SurA binds uOMPs in a groove between the core and P1 domains. We sought to identify the uOMP binding site on SurA by visualizing the high-efficiency *p*AF crosslinking sites on known conformations of SurA (Table 3.4)(69, 150). This analysis revealed that residues colocalized around a groove that forms between the core and P1 domains in the open conformation (Figures 3.1D, 3.12, and 3.13). In this conformation, both of the PPIase domains are structurally isolated from the core domain (shown in Figure 3.1A,B,D).

The uOMP-binding groove is large enough (~25 x 25 x 25 Å) to shield from water the entire length of either a transmembrane β-strand or β-hairpin of an uOMP (Figure 3.13A). The walls of the groove are formed by the core and P1 domains, which provide hydrophobic patches surrounded by weakly positively charged regions. The base of the groove contains a long hydrophobic stretch (30 Å) and is more positively charged than the walls. Interestingly, the regions of the core and P1 domains outside of the groove are highly negatively charged. The separation of charges on SurA could allow electrostatics to play a role in driving uOMP binding to the groove (Figure 3.13B). In sum, the SurA

groove possesses a hybrid chemical nature and size well suited to accommodate the alternating hydrophobic-hydrophilic patterning of uOMP transmembrane domains.

SurA solubilizes uOmpA$_{171}$ in an expanded conformation. uOMPs are expected to exist in a relatively collapsed, molten globule state in the absence of chaperones (153). This collapsed conformation is maintained when uOmpA$_{171}$ is bound by the other major chaperone in the uOMP biogenesis pathway, Skp (111, 128). To determine whether the overall size and shape of a uOMP changes when bound to SurA, we measured the hydrodynamic properties using small angle neutron scattering (SANS). SANS reports directly on the radius of gyration (R_G) and the maximum end-to-end distance (D_{Max}) of macromolecules. Moreover, the sample and buffer conditions can be manipulated in a SANS experiment to visualize a selected component within a complex.

We capitalized on this selective contrast feature of SANS and collected the scattering profile of a photo-crosslinked complex composed of protonated SurA$_{105,pAF}$ and perdeuterated uOmpA$_{171}$ in 30% D_2O (Figure 3.6A). In this experiment, SurA contributes a minor fraction to the scattering contrast (Figure 3.14), and the scattering intensity is primarily contributed by uOmpA$_{171}$. Guinier and P(r) analyses of data collected in this condition revealed that the complex has an R_G value of 45 ± 3 Å and a D_{Max} of 150 ± 10 Å (Figure 3.6B and C and Tables 3.1 and 3.2). These sizes far exceed the expected R_G and D_{Max} calculated from the structure of apo-SurA (R_G = 35 Å; D_{Max} = 105 Å). Previous experiments also show that apo-SurA is not denatured by D_2O (xx). We therefore conclude that the large complex size observed arises from an expanded state of uOmpA$_{171}$ whilst in complex SurA.

To understand the extent to which SurA expands uOmpA$_{171}$ relative to the inherent, unbound uOmpA$_{171}$ molten-globule state, we estimated the intrinsic R_G and D_{Max} of unfolded OmpA$_{171}$. It is impossible to directly measure these parameters with scattering experiments because uOMPs aggregate at the required protein concentrations. However a complementary hydrodynamic parameter of uOmpA$_{171}$, the sedimentation coefficient, s, has been previously reported (s = 1.65 Svedbergs) (173). We connected R_G , D_{Max}, and s-value using HullRad analysis of atomic models of uOmpA$_{171}$ (125). A series of structural models of unfolded OmpA$_{171}$ were created, and the average R_G and D_{Max} of uOmpA$_{171}$ models that agree with the experimental s -value were 25 ± 1 Å and 82.5 ± 9 Å respectively (Figures 3.15 and 3.16; error is estimated from standard deviations). Thus, the R_G and D_{Max} of uOmpA$_{171}$ are both approximately doubled when it is in complex with SurA. In sum, our *p*AF crosslinking and SANS data support a model where client uOMPs are expanded by SurA: a portion of the client uOMP is bound within the SurA groove, and the remainder of the uOMP is poised to sample transient interactions broadly across the SurA surface.

SurA preferentially interacts with specific segments on client uOMPs. The expansion of uOMPs by SurA raises the question of where and how they interact with the groove. To further define the molecular basis of the SurA•uOMP interaction, we used photo-crosslinking mass spectrometry (pXL-MS) to identify the segments on uOMPs that crosslinked to the high efficiency SurA$_{pAF}$ variants. We identified multiple SurA-binding segments on each client uOMP according to the following criteria.

The eight high-efficiency SurA$_{pAF}$ variants were crosslinked to uOmpA$_{171}$ and uOmpX, and subjected to proteolysis with either trypsin only or trypsin and GluC in serial. The resulting peptides were analyzed by LC-MS/MS, and crosslinked peptides were identified (with FDR < 0.01) using the MeroX v2.0 software package (161). Summary data of all pXL-MS experiments are given in Table 3.3 and summary data of all peptide spectrum matches (PSMs) from these pXL-MS experiments are provided as Supplementary Data 1 and 2.

Figure 3.17A shows the crosslinking pattern identified for both uOmpA$_{171}$ and uOmpX for each high efficiency SurA$_{pAF}$ variant. We hypothesized that uOMP residues that repeatedly crosslink to SurA$_{pAF}$ variants (>50%) delineate preferred uOMP binding segments. This strategy was necessary because the high reactivity of the nitrene group formed upon photolysis of *p*AF combined with the difficulty of determining the relative abundance of various crosslinked peptides made it impossible to use the crosslink sites from any individual SurA$_{pAF}$ variant to distinguish preferred binding segments.

Using this criterion, we identified six SurA-binding segments on uOmpA$_{171}$ (residues 1-21, 51-61, 84-86, 95-104, 112-113, and 130-131) and four SurA-binding segments on uOmpX (residues 29-36, 42-48, 64-73, and 81-89) as shown in Figure 3.17A. The identified uOMP binding segments vary in length and location between the client uOMPs and crosslink to residues on both the core and P1 domains of SurA. The sequences of these SurA-binding segments are unusually enriched in tyrosine residues (10 of the 13 tyrosines on uOmpA$_{171}$ appear in segments; P=0.003 by Chi-square test). Indeed SurA has been shown to preferentially bind to peptides with high aromatic content,

affirming our criteria for defining uOMP segments (55). Serendipitously, in OmpA many of these tyrosines are highly conserved according to the Pfam database (Pfam ID: 01389), perhaps indicating importance for these residues in OmpA biogenesis (174).

To validate the importance of the SurA groove as the uOMP binding site, we performed two controls. In the first, we carried out pXL-MS on $SurA_{422,pAF}$, which places the *p*AF away from the groove on the opposite side of the core domain. $SurA_{422,pAF}$ only crosslinks to a single site on $uOmpA_{171}$ and did not crosslink to uOmpX at all, demonstrating a preference of uOMP interactions with the SurA groove (Figure 3.17A). Secondly, we found a similarly small number of crosslinks upon mutation of a highly conserved residue in the construct $SurA_{26,pAF}$ that also resides away from the groove. Taken together, these experiments show that binding segments on uOMPs selectively distinguish and interact with the SurA groove.

Figure 3.17B shows a structural model of $uOmpA_{171}$ in an expanded state with a D_{Max} equal to the experimentally determined size of $uOmpA_{171}$ when bound to SurA. Each of the putative binding segments are highlighted with the same colors used in Figure 3.17A. Two possible SurA•$uOmpA_{171}$ complexes are shown with SurA bound to the second (left) and last two segments (right) of $uOmpA_{171}$. The presence of multiple SurA-binding segments on $uOmpA_{171}$, along with its expanded size, could allow for more than one copy of SurA to simultaneously bind to different segments of a single $uOmpA_{171}$ (Figure 3.17B), as further supported by the presence of higher molecular weight bands in SDS-PAGE of some crosslinked $SurA_{pAF}$ variants (Figure 3.3).

Modeling the Structural Features of SurA•uOmpA$_{171}$. Our SDS-PAGE (Figures 3.1B & 3.3), SANS (Figure 3.6), and pXL-MS (Figure 3.17) experiments support a mechanism in which SurA binds a defined client uOMP segment in the SurA groove. The remainder of the uOMP is greatly expanded and presumed to be dynamic. To visualize this binding mode in more detail, we built 23 models of uOmpA$_{171}$ bound to SurA using our experimental findings as restraints.

Figures 3.9 and 3.10 provide flowcharts describing the process by which these models were generated and where each piece of experimental information was included. In essence, we docked the binding segments of uOmpA$_{171}$ identified by pXL-MS to the groove of SurA using distance restraints generated by HADDOCK (166). The uOmpA$_{171}$ component of these models were then expanded to be compatible with the SANS data (Figure 3.18A and B).

The Hill coefficient reported for SurA binding uOmpA$_{171}$ is greater than 1, indicating more than one copy of SurA can interact with uOmpA$_{171}$ at a time (146). This multiplicity is also consistent with our finding of higher molecular weight bands in SurA$_{pAF}$ SDS-PAGE experiments. Accordingly, we also created 17 models with additional SurA protomers docked to the expanded uOmpA$_{171}$ (Figure 3.18 C-E). Hydrodynamic parameters of each of the 40 structural models created are tabulated in Table 3.5.

Chemical crosslinking validates features of SurA•uOmpA$_{171}$ models, and reveals distinct binding modes. To validate structural features of the SurA•uOmpA$_{171}$ sparse ensemble, we performed XL-MS with the chemical crosslinker disuccinimidyl dibutyric urea (DSBU) on WT SurA and our client uOMPs. In total, we identified 46 unique DSBU

crosslinks between SurA and uOmpA$_{171}$ and 17 unique DSBU crosslinks between SurA and uOmpX (Figure 3.19 and Table 3.3, and Supplementary Data 3–5).

To ascertain underlying similarities in our structural models, we performed spectral biclustering on a matrix of all solvent-accessible surface distances (SASDs, calculated using JWalk) for the 46 identified crosslinks in all 23 structural models of SurA•uOmpA$_{171}$ (Figure 3.8 shows this analysis, Supplementary Data 6 provides all SASDs and their associated scores) (175–177). Due to the conformational heterogeneity and multiplicity of the SurA•uOmpA$_{171}$ sparse ensemble, no single structural model captured all of the DSBU crosslinks. However, three distinct SurA•uOmpA$_{171}$ binding modes emerged from this analysis, which are each explained by a unique subset of the identified crosslinks.

Figure 3.20A shows an example of the first binding mode, which is found in seven of our SurA•uOmpA$_{171}$ models, wherein segment 1 is bound in the groove and the remainder of uOmpA$_{171}$ is projecting away from SurA. The second binding mode, shown in Figure 3.20B, is present in five SurA•uOmpA$_{171}$ models and is similar to binding mode one, except segment 2 is bound in the groove. Seven members of the sparse ensemble evinced a more complex topology in which the uOMP threads through the groove several times, defining a third binding mode (Figure 3.20C). Similar to the first binding mode, segment 1 is bound in the groove, but in this latter mode segments 3, 4, and 5 now make extensive contacts with the groove and P1 on a “second pass.” Together, these three binding modes of the SurA•uOmpA$_{171}$ complex cover 75% (34 out of 46; Figure 3.8) of the identified DSBU crosslinks and help explain why so many regions of uOmpA$_{171}$ can crosslink with the SurA groove (Figure 3.17).

As described above, we constructed (SurA)$_n$•uOmpA$_{171}$ models with higher-order stoichiometries (where $n = \{2, 3, 4\}$). These models provide an additional explanation for the complex DSBU crosslinking pattern between uOmpA$_{171}$ and SurA. Figure 3.20D shows how a single uOmpA$_{171}$ can distribute these binding modes across more than one copy of SurA instead of threading itself through a single copy of SurA multiple times – accommodating the same clusters of crosslinks. Moreover, inclusion of models with higher order stoichiometries increases coverage to 89% of identified DSBU crosslinks (Supplementary Data 6).

Finally, as a critical control to confirm the SurA groove as the primary binding site for uOMP substrates, we performed DSBU crosslinking experiments between uOmpA$_{171}$ and the "locked-closed" SurA variant (P61C/A218C) previously described by Silhavy and co-workers (40). In this variant the open conformation of SurA is inaccessible, and cells show increased sensitivity to viability envelope stressors *in vivo*. We observed a drastic reduction in the number of "locked-closed" SurA–uOmpA$_{171}$ interprotein crosslinks compared to WT SurA (7 vs 46; Figure 3.19, Supplementary Data 3 & 5). This finding implies that the dense crosslinking patterns revealed by XL-MS depend on the formation of the SurA groove, and not from contacts made outside of the groove.

SANS validates and quantifies XL-MS based binding modes. To test which SurA•uOmpA$_{171}$ models are representative of the conformational ensemble in solution, we compared them to an independent SANS profile of SurA$_{105,pAF}$-uOmpA$_{171}$ collected in 0% D_2O (Figure 3.7). In this particular condition, protonated SurA and perdeuterated uOmpA$_{171}$ contribute equally to scattering (Figure 3.14). We calculated the expected

scattering profile that would arise under these experimental conditions for each of the 40 SurA•uOmpA$_{171}$ models using the software package, SASSIE (113). In addition to the models of the SurA•uOmpA$_{171}$ complex, we included multiple models of apo-SurA in varying conformations as we were unable to completely purify the crosslinked complex (Figure 3.5). In agreement with the DSBU crosslinking, we found that no single model, or pairs of models, recapitulated the experimental scattering profile (reduced chi-square < 1.05) (116).

Linear combinations of the scattering profiles from three structural models were able to describe the SANS dataset. In total, we sampled over 1 million combinations of triplets of models and found 35 combinations whose simulated SANS profiles produced reduced chi-square values less than 1.05 with respect to the experimentally observed 0% D_2O SANS profile. Each accepted triplet contained at least one model of non-crosslinked SurA and one SurA•uOmpA$_{171}$ complex (Table 3.8). The three models most likely to be included in an accepted combination arise from the three distinct binding modes defined by XL-MS, and are shown in Figure 3.20A-C with their associated crosslinks. In addition, the linear combinations identify several models with higher-order stoichiometries as depicted in Figure 3.20D.

The sparse ensemble illustrating the main structural features identified by our experiments is depicted in Figure 3.21 by an overlay of SurA•uOmpA$_{171}$ models (population of each model in the ensemble are listed in Table 3.9). We note that this collection of conformations captures all known experimental data on SurA•uOMP complexes: including uOMP binding to the SurA groove (Figure 3.1), the Hill coefficient

that is slightly greater than 1, the expansion of uOMPs by SurA (Figure 3.6), pXL-MS identification of specific binding segments (Figure 3.16), and the independent validation of binding modes by SANS and XL-MS (Figure 3.20 and 3.21) (40, 57, 60, 69). This sparse ensemble of models of the SurA•uOmpA$_{171}$ complex provides a chemically reasonable and minimalist set of structures that could exist in the full conformational ensemble of the complex in solution. Due to limitations in the resolution of the data, additional conformations that fit the data likely exist.

3.4 Discussion

The periplasmic chaperone network is integral for *E. coli* outer membrane protein biogenesis. SurA plays several important roles in the uOMP biogenesis pathway: it must (i) recognize uOMP clients before they aggregate; (ii) maintain them in a soluble form in the periplasm; and (iii) mediate a hand-off to the BAM complex. In this study we have utilized an integrative/hybrid structural biology approach that combines multiple crosslinking and scattering techniques to generate restraints used to build a representative ensemble. This ensemble captures key structural features of SurA in complex with its client uOmpA$_{171}$.

The model-independent Guinier analysis of the contrast-matched SANS experiments reveal that SurA performs its functions by dramatically expanding the SurA-uOMP complex in a mechanism reminiscent of trigger factor, a structural homolog to SurA (178). We observe a primary uOMP interaction in the SurA groove, located between the core and P1 domains which recapitulates recently published findings (69). The remainder of the unfolded uOmpA$_{171}$ chain that is not occupying the groove must assume an elongated conformation to be consistent with these SANS data. In contrast, a covalently "locked-closed" variant of SurA is unable to efficiently interact with uOmpA$_{171}$, consistent with the reduced functionality of this variant *in vivo.* All together these results suggest uOMP binding to the SurA groove results in uOMP expansion relative to its compact native state, as exemplified in Figures 3.17, 3.20, and 3.21.

This expansion of uOMPs is distinctive from a recently published model of the SurA-uOmpX complex built from crosslinking data alone (69). In that model, a single SurA

completely encapsulates a globular uOMP, and this structural interpretation is incongruous with our experimental findings as the present SANS data demonstrate that the uOmpA$_{171}$ must instead be expanded. Although the reported crosslinking is consistent with our data, our approach also capitalized upon the usefulness of a hydrodynamic view of highly dynamic structures. Instead, our data for the SurA-uOmpA$_{171}$ complex are more consistent with a dynamic conformational ensemble proposed for the SurA-FhuA complex based on NMR (60). The extent of expansion of unfolded FhuA is unresolved, however, because the hydrodynamic properties of this system have not yet been established.

Unfolded OMP expansion mediated by a single SurA binding event is aided by the ability of additional copies of SurA to interact with distal binding segments on the uOMP (as shown in Figures 3.17B and 3.20D). Expansion of uOMPs would help avoid steric clashes between different copies of SurA simultaneously interacting with an uOMP client. For the small, eight-stranded uOMPs investigated here, higher-order stoichiometries represent minor populations in the ensemble. It is reasonable to speculate that the length of a client uOMP could dictate the binding stoichiometry of the SurA•uOMP complex as larger clients would ostensibly contain a greater number of SurA-binding segments. In accordance with this idea, gel filtration data suggest that the stoichiometry for the SurA-FhuA complex may be closer to 2:1 because FhuA is a much larger outer membrane protein (22 β-strands as compared to 8 for OmpA) (56, 60). Higher-order stoichiometries could also be enhanced in the crowded periplasm, where the excluded volume effect increases protein-protein interactions (179).

The initial finding of uOmpA$_{171}$ expansion was surprising, especially because the other major periplasmic chaperone Skp encapsulates uOMPs in a collapsed state reminiscent of their intrinsic, unfolded conformation (60, 111). Given SurA interacts with uOMPs primarily through its groove, the persistent global expansion of the remainder of the client uOMP at first glance appears puzzling. We propose a kinetic trapping mechanism wherein binding and release of uOMP segments are fast relative to the collapse of the uOMP to its intrinsic molten globule state. Indeed, kinetic partitioning is a dominant organizing feature of the uOMP biogenesis chaperone network, and the interaction between SurA and uOMP happens on a very fast time scale (47, 61, 180). Rates of unfolded uOMP collapse are not known but may be relatively slow given the low overall hydrophobicity of transmembrane β-barrel primary sequences. Such a difference in the rates of uOMP intrinsic collapse versus expansion by SurA binding would provide a way to retain uOMPs in an expanded state through transient repeated associations with SurA protomers. This mechanism has the advantage of limiting the amount of SurA required to solubilize a client thereby maximizing the reservoir of free SurA in the periplasm.

Figure 3.22 highlights several biological implications of the features of our structural ensemble. We hypothesize that these SurA properties could explain its multifaceted roles in OMP biogenesis. Firstly, we expect that expansion of uOMPs would decrease unproductive intraprotein interactions and maintain the chain in a folding-competent, unfolded conformation (Figure 3.22B). In this respect, SurA performs an orthogonal role to the other chaperones in the uOMP biogenesis network, which form

cages around uOMPs to decrease interprotein interactions. Expansion may also mediate the formation of transient hetero-chaperone complexes in the periplasm, where more than one chaperones are simultaneously bound to an uOMP (Figure 3.22C and D). As uOMPs are expected to undergo hundreds individual binding and dissociation events while in the periplasm, a SurA-mediated hand-off of expanded uOMPs between chaperones would allow for these events to occur while keeping the population of aggregation-prone, unbound uOMP low (47).

An intriguing outcome of our structural ensemble is the finding that SurA-binding segments of client uOMPs are located toward the N-terminus (Figure 3.17). Indeed, the three binding modes found for the SurA•uOmpA$_{171}$ interaction are mediated by the two most N-terminal binding segments on uOmpA$_{171}$. This suggests the possibility of an early encounter between SurA and client uOMPs in the periplasm (Figure 3.22A). Accordingly, a recent, low resolution cryo-EM model places SurA near the translocon where it is well positioned to bind emerging uOMP segments (131).

Conversely, there was a conspicuous absence of robust crosslinking for both client OMPs near their C-termini. The apparent lack of SurA interaction sites in this uOMP region leaves the β-signal (Aro-X-Aro) free to interact with other members of the uOMP biogenesis pathway. The β-signal has been shown to play an important role in mediating efficient catalysis of uOMP folding by BAM both *in vivo* and *in vitro* (181–183). As SurA is the only periplasmic chaperone that promotes the interaction between uOMPs and BAM, our data support a mechanistic hypothesis in which this region of uOMPs is free

and flexible and effectively ‘cast’ outward in a mechanism reminiscent of fly fishing to catch the BAM complex (Figure 3.22E) (66, 184).

Even if it is transient, the formation of a SurA•uOMP•BAM ternary complex is enticing because it brings the uOMP close to both the BAM complex and the disrupted adjacent membrane, both of which accelerate uOMP folding (24, 42). The BAM complex has been proposed to template and insert uOMP β-hairpins as individual foldamers of OMPs. Notably, the SurA groove is large enough to accommodate a β-hairpin, which could potentially favor the pre-formation of this key structural elements in a nascent uOMP. Moreover, this SurA-mediated β-hairpin formation mechanism could be easily adapted to larger clients with more transmembrane strands (and probably more SurA-binding segments) given the modular nature of the β-hairpin unit.

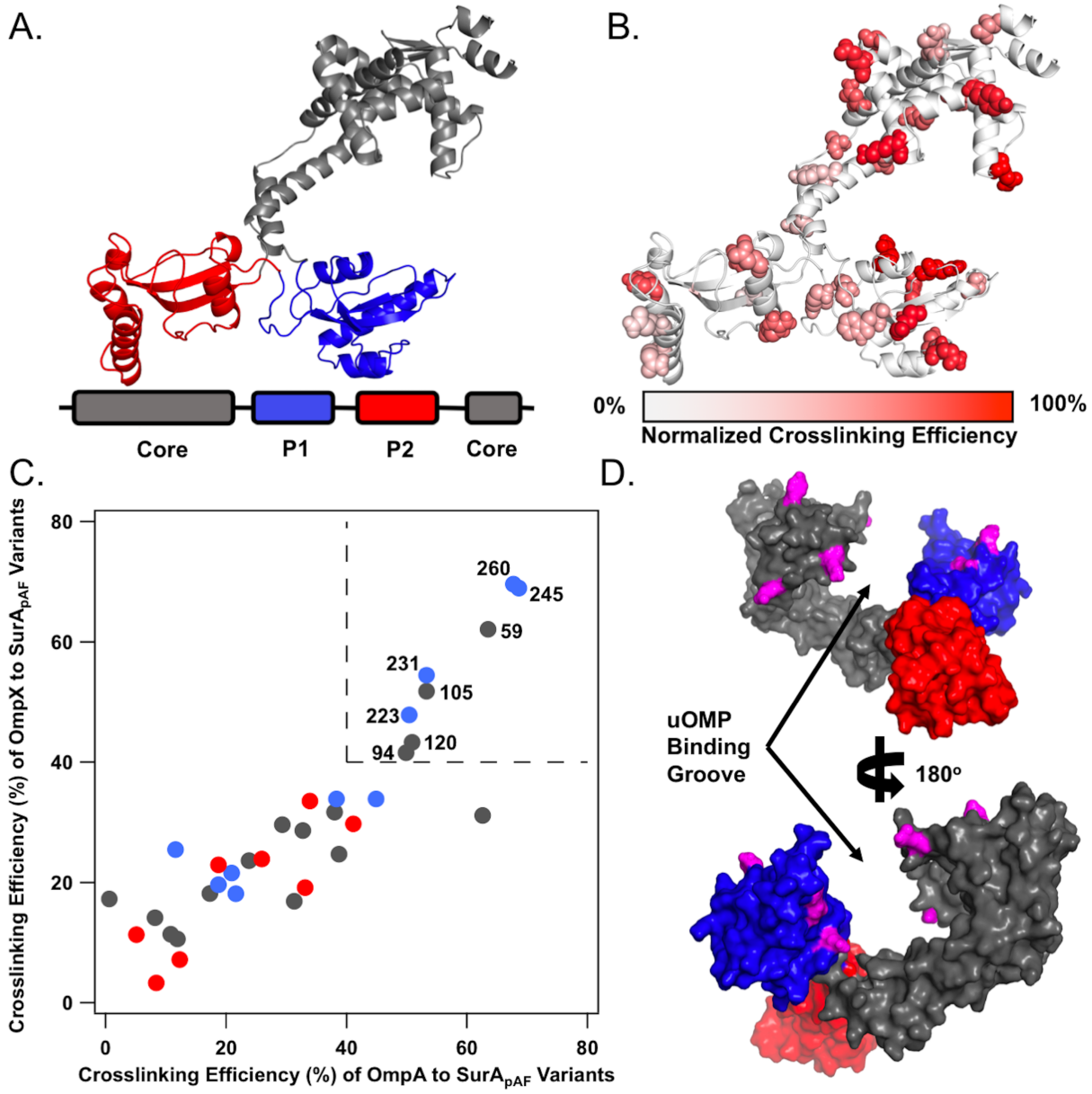
In this work we highlight the utility and complementarity of photo- and chemical crosslinking, neutron scattering, and mass spectrometry applied together. This combined approach was crucial because unfolded proteins present many challenges to conventional structural techniques due of their absence of regular secondary structure, their high conformational flexibility, and their propensity to aggregate. Our results illuminate a sparse ensemble of models that captures the key structural features defining how SurA promotes outer membrane protein biogenesis.

3.5 Acknowledgements

We thank the Johns Hopkins University Biomolecular NMR Center and the Center for Molecular Biophysics for providing facilities and resources. Access to NGB30 SANS was provided by the Center for High Resolution Neutron Scattering, a partnership between the National Institute of Standards and Technology and the National Science Foundation under Agreement No. DMR-1508249. We acknowledge the support of the National Institute of Standards and Technology, U.S. Department of Commerce, in providing the neutron research facilities used in this work. This work benefitted from CCP-SAS software developed through a joint EPSRC (EP/K039121/1) and NSF (CHE-1265821) grant. This work was supported by National Science Foundation (NSF) grant MCB1412108. D.C.M, A.M.P., T.D. and H.J.L. were supported by NIH training grant T32 GM008403. A.M.P. was supported by the NSF grant DGE 1232825. The authors thank lab members for helpful discussions.

3.6 Figures

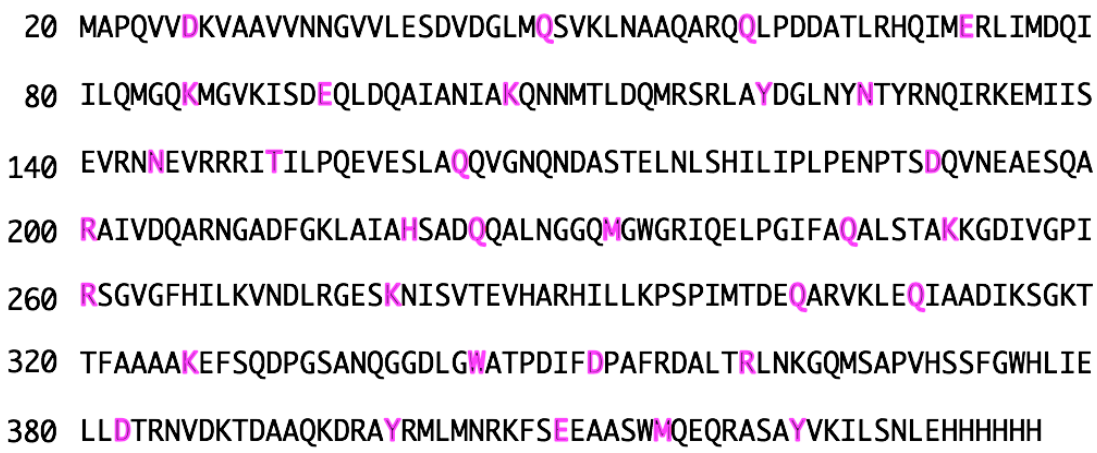
Figure 3.1 - Open SurA binds client uOMPs in a groove between domains.



(A) The structure of open SurA shown as a cartoon, with the domains colored as depicted in the sequence diagram below (core: gray, P1: blue, P2: red). In this conformation, the three domains of SurA are structurally isolated from each other and do not form extensive inter-domain contacts. **(B)** The 32 surface exposed sites on SurA in which *p*AF was

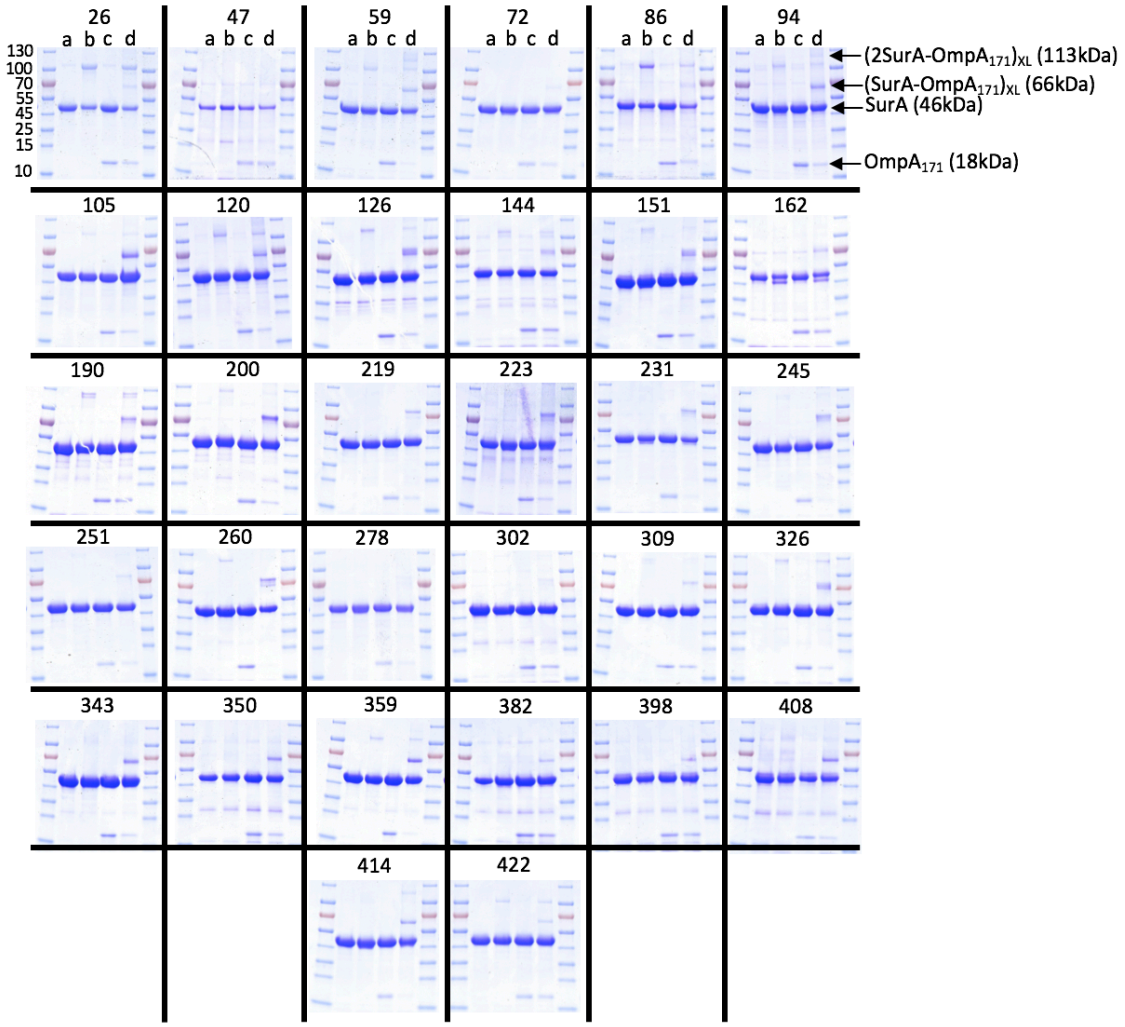
substituted, shown in a space-filling representation. Photo-crosslinking was induced with 5-min UV exposure. Each crosslinking site is colored based on the normalized crosslinking efficiency to $uOmpA_{171}$ as based on quantitative SDS-PAGE (Figure 3.3). The highest crosslinking sites are found on the core and P1 domains of SurA, while P2 exhibits only modest crosslinking efficiency. (C) The raw crosslinking efficiencies of SurA to $uOmpA_{171}$ and uOmpX are shown and colored by the SurA domain in which they residue (as in **(A)**). Eight $SurA_{pAF}$ variants stand out by having high (>40%) crosslinking efficiency to both uOMP clients, and are labeled with their residue number in the upper right quadrant of the graph (demarcated by dotted lines). **(D)** The eight high efficiency crosslinking sites, shown in magenta, are mapped on to a surface representation of the structure of open SurA. Together, these sites line a groove formed between the core and P1 domains, indicating that uOMPs are primarily bound there.

Figure 3.2 - SurA sequence with pAF sites highlighted.



The sequence of SurA is shown, with the residues replaced by para-azido phenylalanine (*p*AF) highlighted in magenta. These sites were chosen on the basis that they are surface exposed in the monomeric crystal structure of SurA (PDB: 1M5Y) and are polar in nature to minimize the effects of pAF incorporation on protein structure and function.

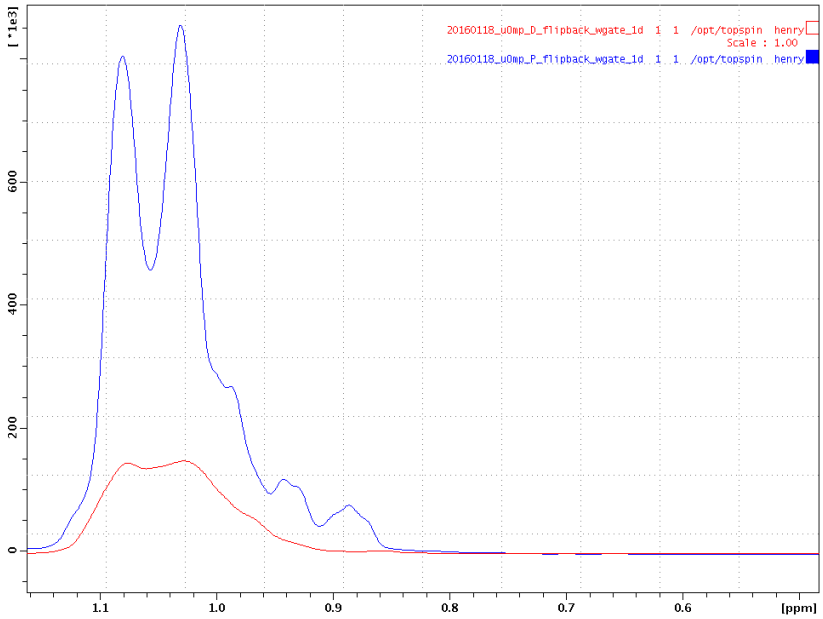
Figure 3.3 - Crosslinking experiments suggest that SurA binds to uOmpA$_{171}$ with a delocalized interface.



SurA$_{pAF}$ (25 μM) with or without uOmpA$_{171}$ (5 μM) was reconstituted in 20 mM Tris (pH 8), and exposed (or not) to UV light for 5 min. The resulting photo-products were then resolved and analyzed by SDS-PAGE. Representative SDS-PAGE analysis for crosslinking experiments between 36 SurA$_{pAF}$ variants and uOmpA$_{171}$ are shown. For each gel, the lanes are loaded as follows: SurA alone (**a** (-UV) and **b** (+UV)); SurA + uOmpA$_{171}$ mixture (**c** (-UV) and **d** (+UV)). Prior to UV exposure, the SurA variants and

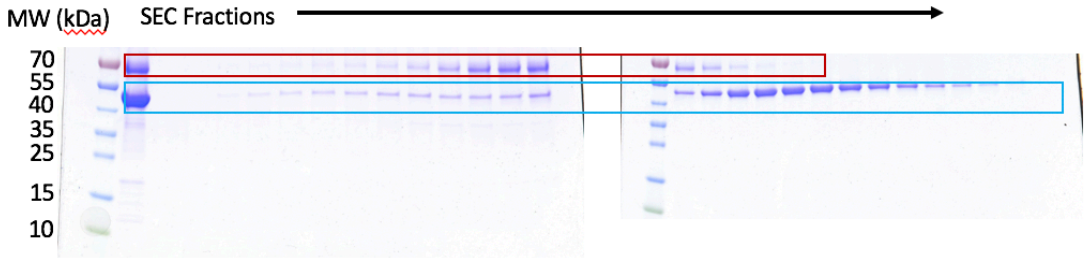
uOmpA$_{171}$ are observed as bands with apparent molecular weights of 46 kDa and 18 kDa, respectively. After UV exposure, some variants show a higher apparent molecular weight band (*i.e.*, “Complex”). The migration positions of the one-to-one and two-to-one species are shown for SurA$_{94,pAF}$.

Figure 3.4 - NMR Determination of Deuteration Level of uOmpA used in some SANS Experiments.



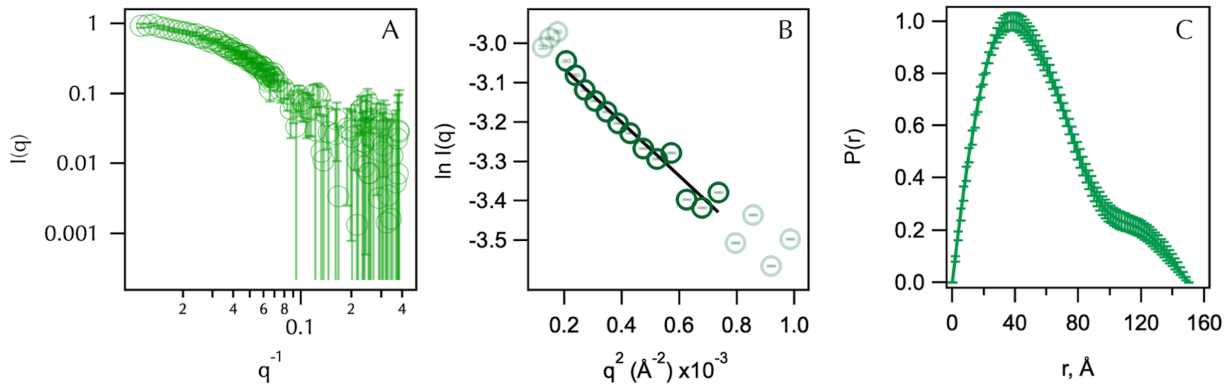
OMP deuteration was estimated to be 80%. We collected 1D ^{1}H spectra on both protonated and deuterated $uOmpA_{171}$ (50 mM) in 8 M Urea, 20 mM Tris, 10% D_2O at 35 °C, on a 600 MHz Bruker Avance II spectrometer. Water suppression was achieved using a flipback-watergate sequence and a buffer purging pulse was included to minimize the large urea peak. Each spectrum was collected using 128 scans, a recycle delay of 1.5 s, and acquisition time of 60 ms/FID. Data was processed and analyzed using TopSpin 2.1. The spectra were aligned using the amide resonance peaks, and the baseline of the methyl peaks was corrected using a 5^{th} order polynomial. After baseline correction, the methyl peak volumes were integrated using TopSpin, and the integrated intensities of the methyl peaks in both the protonated and deuterated samples were compared. The loss of intensity in the methyl peaks between the two samples was used to estimate the deuteration level of deuterated-$uOmpA_{171}$.

Figure 3.5 - SDS PAGE of SEC Fractions indicates crosslinked SurA-uOmpA$_{171}$ cannot be fully separated from excess SurA.



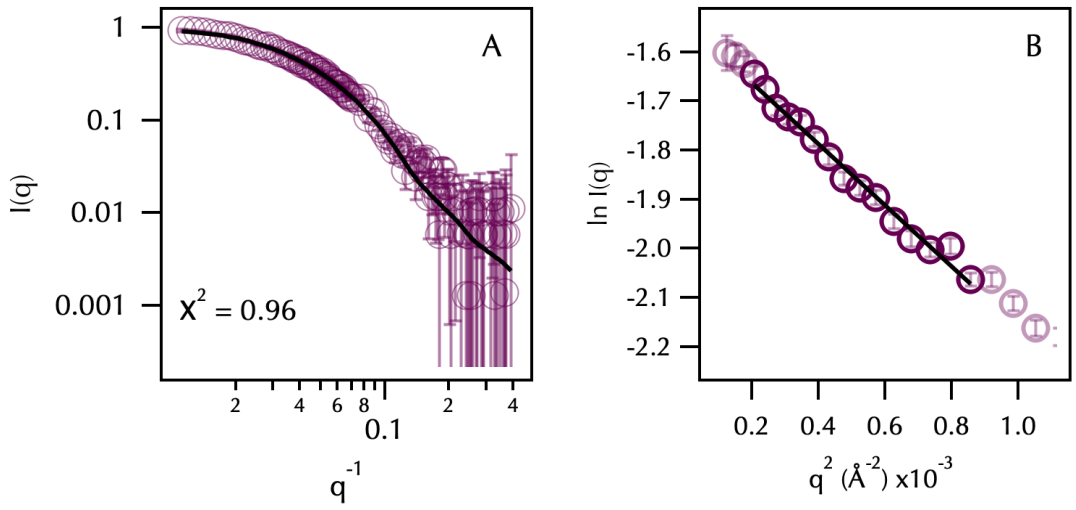
SDS PAGE gel shows the crosslinked SurA$_{105,pAF}$-uOmpA171 sample that was injected on to the SEC column (first lane on left gel). Further lanes are 0.5mL fractions collected from the SEC run, with the crosslinked complex (red box) and free SurA (blue box) highlighted. Because of the small change in size (45 vs 65 kDa), and the possibility that free SurA is interacting with the crosslinked complex, we could not completely separate the two species using SEC.

Figure 3.6 - SANS of a SurA$_{105,pAF}$-uOmpA$_{171}$ complex reveals an expanded uOmpA$_{171}$.



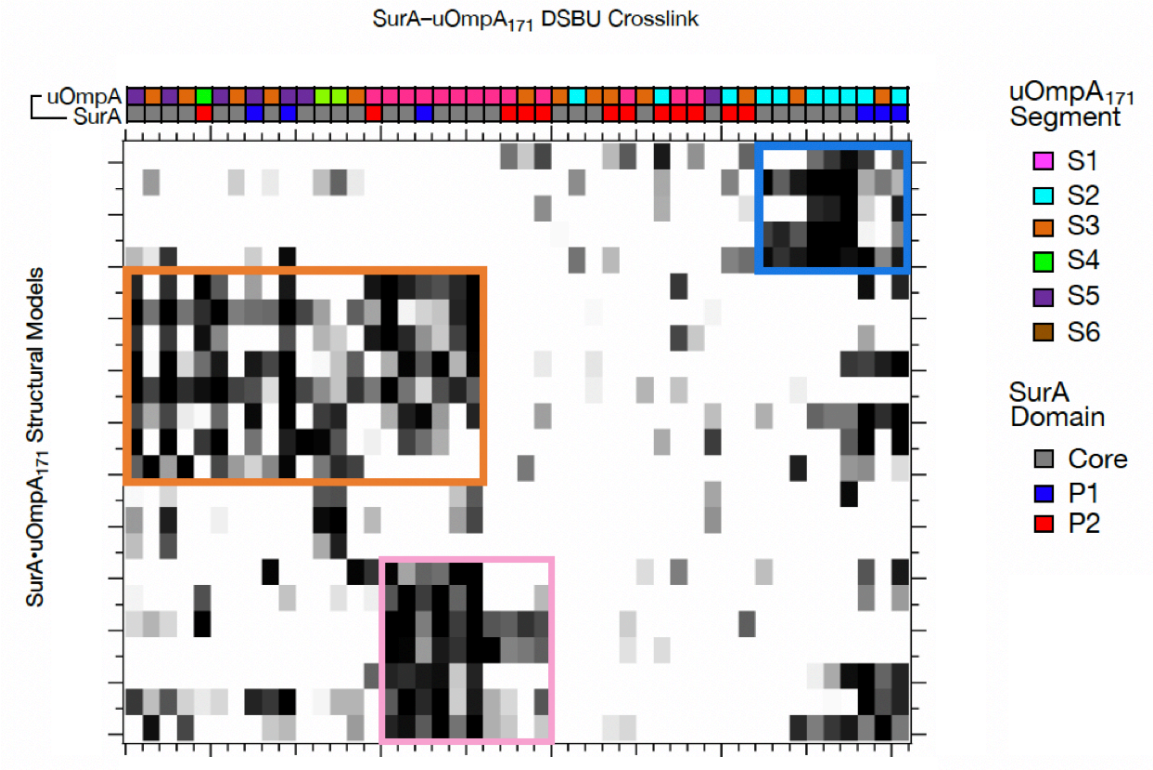
(**A**) Raw scattering profile of protonated-SurA$_{105,pAF}$ photo-crosslinked to deuterated-uOmpA$_{171}$ in 30% D_2O buffer is shown in green. Error bars represent the standard error of the mean with respect to the number of pixels used in the data averaging. (**B**) Linear fit of the Guinier region of the SANS profile determines the R_G of the complex to be 45 ± 3 Å. (**C**) $P(R)$ distribution function inferred from the Guinier fit; D_{Max} is estimated to be 150 ± 10 Å.

Figure 3.7 - SANS Profile of $SurA_{105,pAF}$-$uOmpA_{171}$ crosslinked complex in 0% D_2O



(**A**) SANS data for $(SurA_{105,pAF}–uOmpA_{171})_{XL}$ with SurA protonated and $uOmpA_{171}$ perdeuterated, for the 0% D_2O buffer condition. The black lines through the data are the average waves from triplet basis setting with the χ^2 for the fits. (**B**) Guinier regions of these data with a linear fit (values for the Guinier fits Table 3.1).

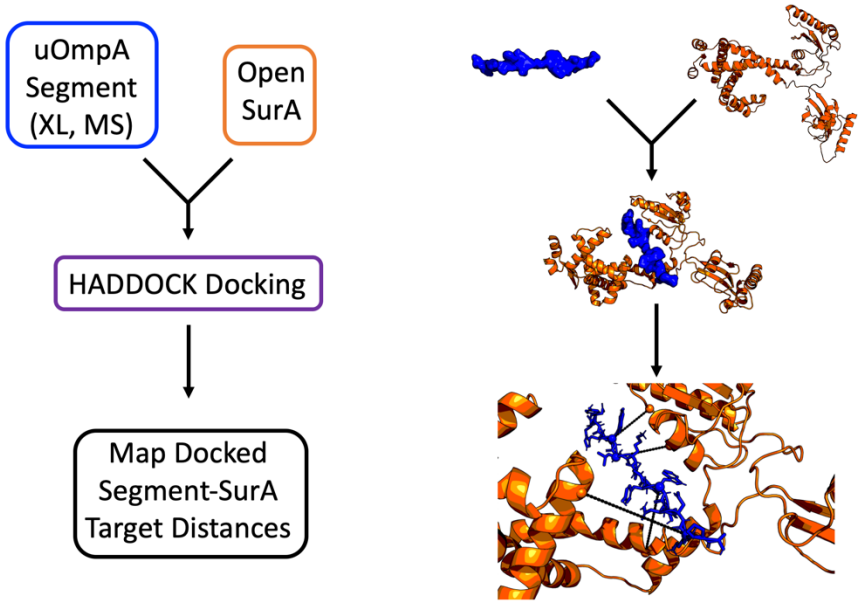
Figure 3.8 - Clusters among DSBU crosslinks and three SurA·uOmpA$_{171}$ binding modes



Spectral biclustering reveals a natural grouping among structural models (rows) that are collectively consistent with subsets of crosslinks (columns) (177). Structures divide into a pink cluster (wherein SurA binds segment 1; 7 models), a blue cluster (wherein SurA binds segment 2; 5 models), and an orange cluster (wherein SurA binds uOmpA at segment 1 and either segments 3–5; 8 models). A few of the structures are not well explained by any of the crosslinks. Crosslinks divide into a pink cluster (which support the pink structures, and to a lesser extent, the orange structures), a blue cluster (which support the blue structures), and an orange cluster (which support the orange structures). The blue cluster has a sub-cluster (blue-pink) which is also consistent with some of the pink structures. Twelve of the crosslinks are not well explained by any of the 23

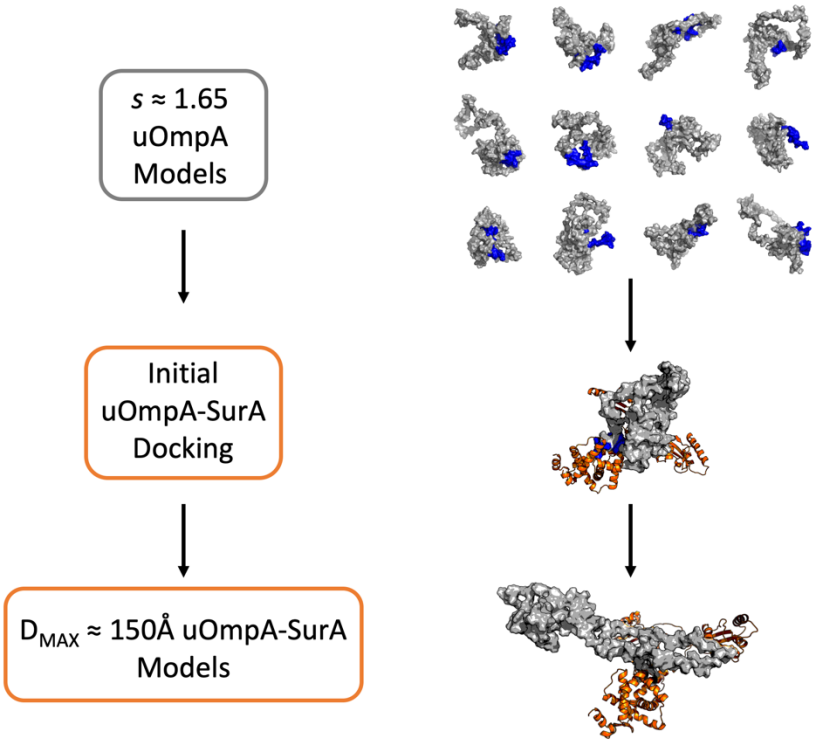
SurA•uOmpA$_{171}$ models. Crosslinks are annotated with colors, with the SurA site represented by the domain it is on and the uOmpA$_{171}$ site represented by the nearest binding segment. The grayscale of the matrix represents the scores for a given crosslink in a given structure (white = 0, black = 100; see SI methods for explanation); all SASDs and scores are in Supplementary Data 6.

Figure 3.9 - HADDOCK docking of uOmpA$_{171}$ segments to SurA



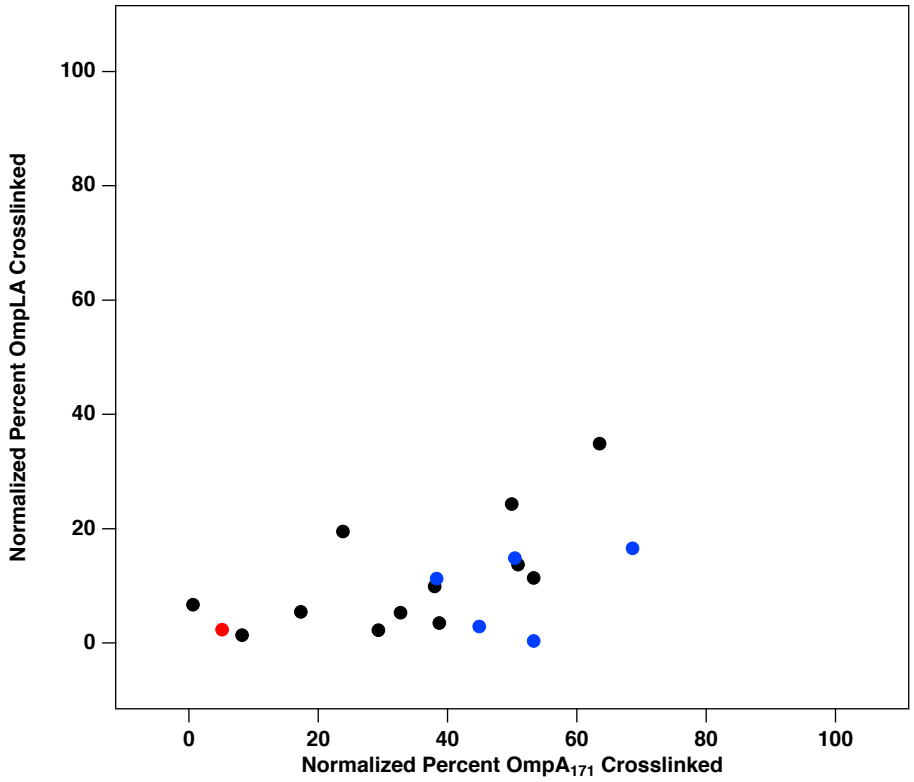
Four OmpA$_{171}$ amino acid segments (2-21, 54-73, 84-104, 115-132) which contain all of the SurA binding segments identified with XL-MS were modelled as extended polypeptides (ϕ=-78˚, ψ=149˚, blue spheres, top right panel). These segments were individually docked to the open form of SurA (orange ribbon) using HADDOCK (166). High ranking docked segment-SurA complexes were inspected to obtain target distances between adjacent uOmpA segment and SurA residues (bottom right panel). These target distances were used to dock full length uOmpA$_{171}$ models to “open” SurA (Figure 3.10).

Figure 3.10 - Docking $uOmpA_{171}$ to SurA and Expanding Bound $uOmpA_{171}$ Flowchart



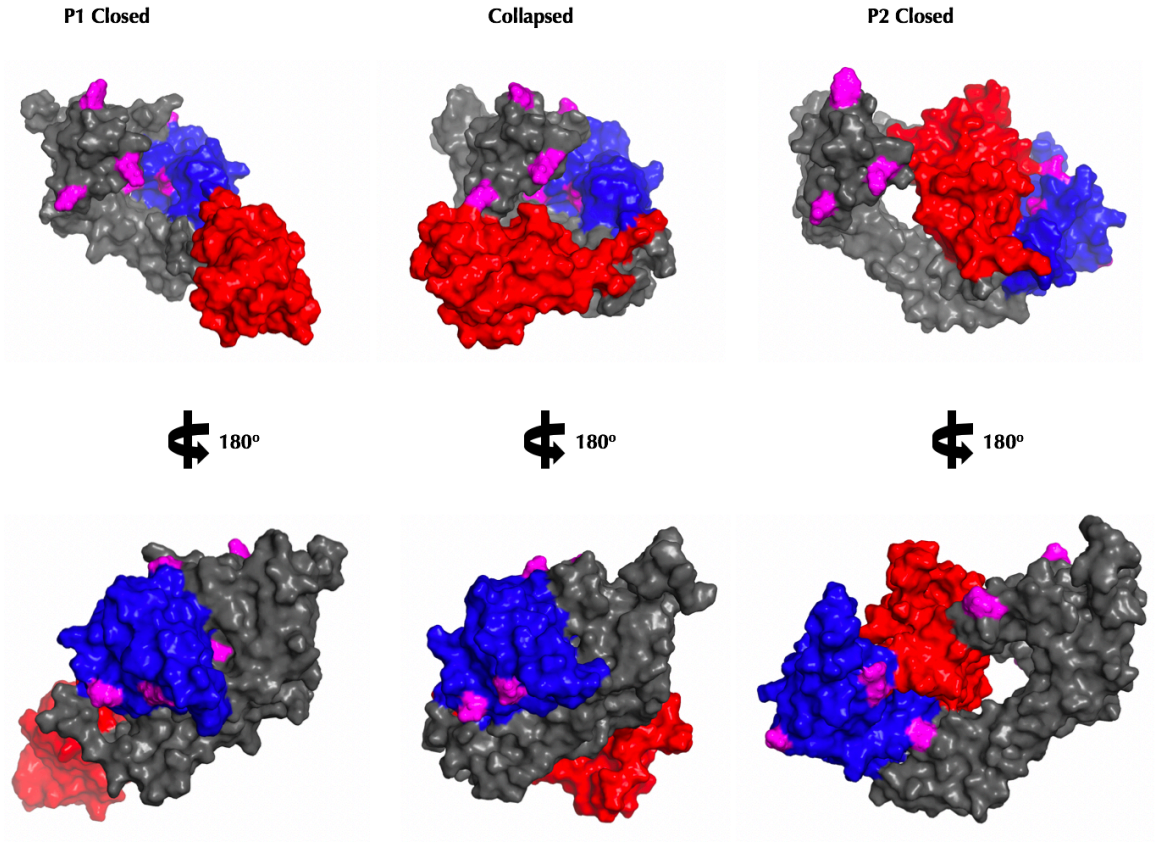
Individual uOmpA models with the corresponding crosslinking segment on the protein surface (blue, top right panel) were identified and docked to open form SurA using the target distances obtained from HADDOCK docking (middle panel). The collective variables module in NAMD with implicit solvent conditions was used to remove atomic clash during docking. Additional uOmpA-SurA models were made by expanding the maximum dimensions of docked uOmpA to ~150 Å as suggested by SANS analysis in 30% D_2O (lower panel).

Figure 3.11 - Crosslinking Efficiency of the Non-Cognate Client OmpLA is Low



Normalized crosslinking efficiencies of the 36 $SurA_{pAF}$ variants to $uOmpA_{171}$ and uOmpLA. Each point represents an individual $SurA_{pAF}$ variant (colored by the domain of which it is part). The crosslinking efficiencies of SurA to uOmpLA were much lower and do not correlate to their efficiencies to $uOmpA_{171}$, consistent with a hypothesis that uOmpLA is not a substrate for SurA and crosslinks to it reflect non-specific associations.

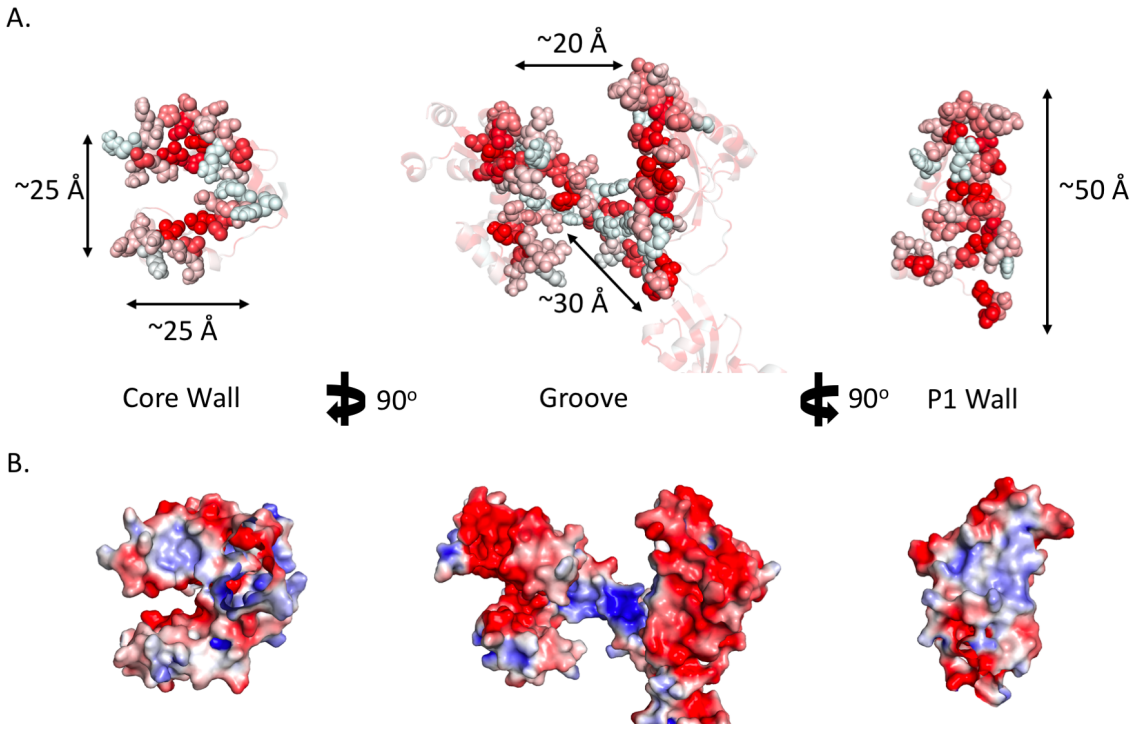
Figure 3.12 - Compact apo SurA structures do not colocalize high-efficiency crosslinking sites



SurA is shown in a surface representation with domains in each model colored as in Figure 3.1A. The positions of the eight high-efficiency crosslinking sites are shown in magenta. In the P1 closed conformation (identified in x-ray crystallography), the pink residues found on the core and P1 domains are on opposite sides of the protein (shown by 180° rotation), which does not allow for a distinct uOMP interaction site to be identified. The collapsed conformation (where both P1 and P2 domains are bound to the core domain) and the P2-closed conformation (where P2 is bound to the core domain and P1 is structurally isolated) have both been recently shown to possibly exist in solution (69).

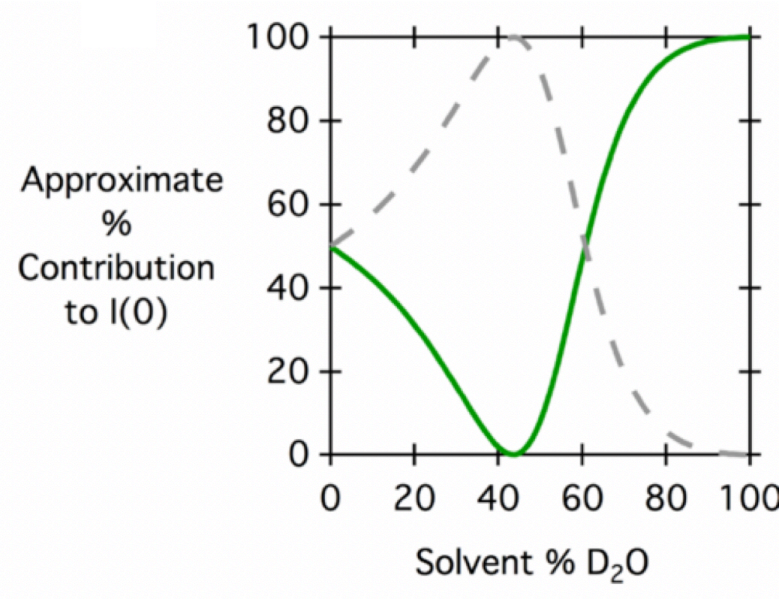
Models created of these conformations also did not allow for a distinct uOMP interaction site to be identified.

Figure 3.13 - Structural Analysis of uOMP Binding Groove in "open" SurA



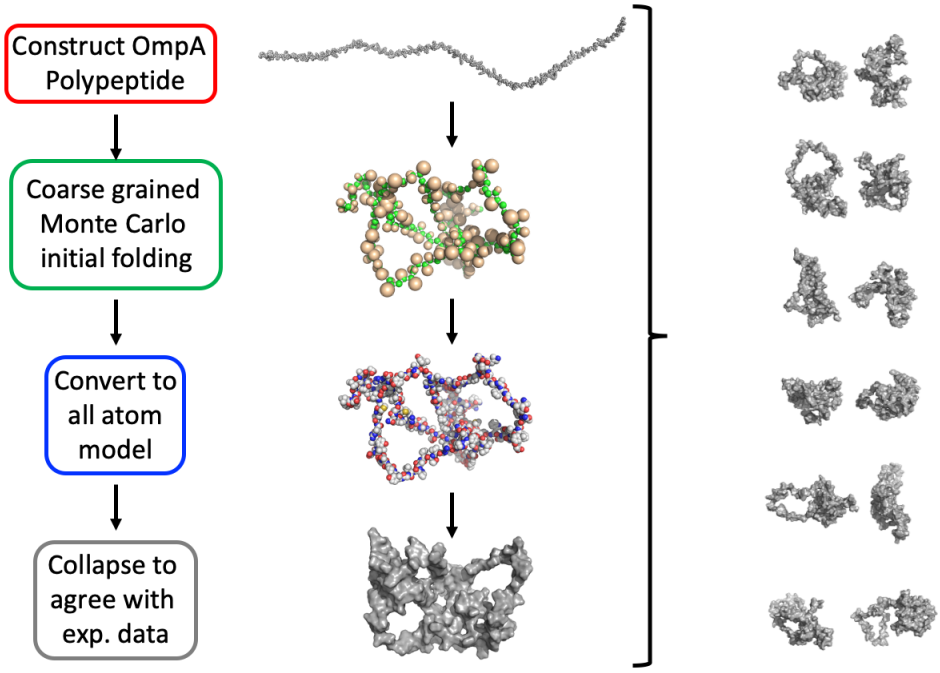
In the "open" conformation of SurA, a groove forms that contains hydrophobic patches and is electropositive in nature. In the middle of each panel, the groove is shown from a top-down perspective; to the left and right are 90˚ rotations, illustrating the contributions of the core domain and the P1 domain to create the "walls" of the groove. Panel **A** shows the groove in a space-filling representation, with residues colored based on hydrophobicity (red is most hydrophobic). The dimensions of the groove are also denoted in this panel, including the length of the floor of the groove, which is made of the C-terminal helix of the core domain. Panel **B** show the groove in a surface representation colored based on the electrostatic potential of the surface (± 3k/T). The floor and walls of the groove are positively charged (blue), while the surface near the top of the groove is negatively charged (red).

Figure 3.14 - Scattering Contribution for SurA and uOmpA$_{171}$ as a Function of %D$_2$O



The contributions to the overall scattering intensity derived from protonated-SurA (green, solid line) and perdeuterated-uOmpA$_{171}$ (gray, dashed line) are plotted as a function of percent D$_2$O in the sample buffer. In our SANS experiments, we utilize two conditions 0% D$_2$O (where each protein contributes equally to scattering) and 30% D$_2$O (where uOmpA$_{171}$ contributes 84% of the scattering intensity).

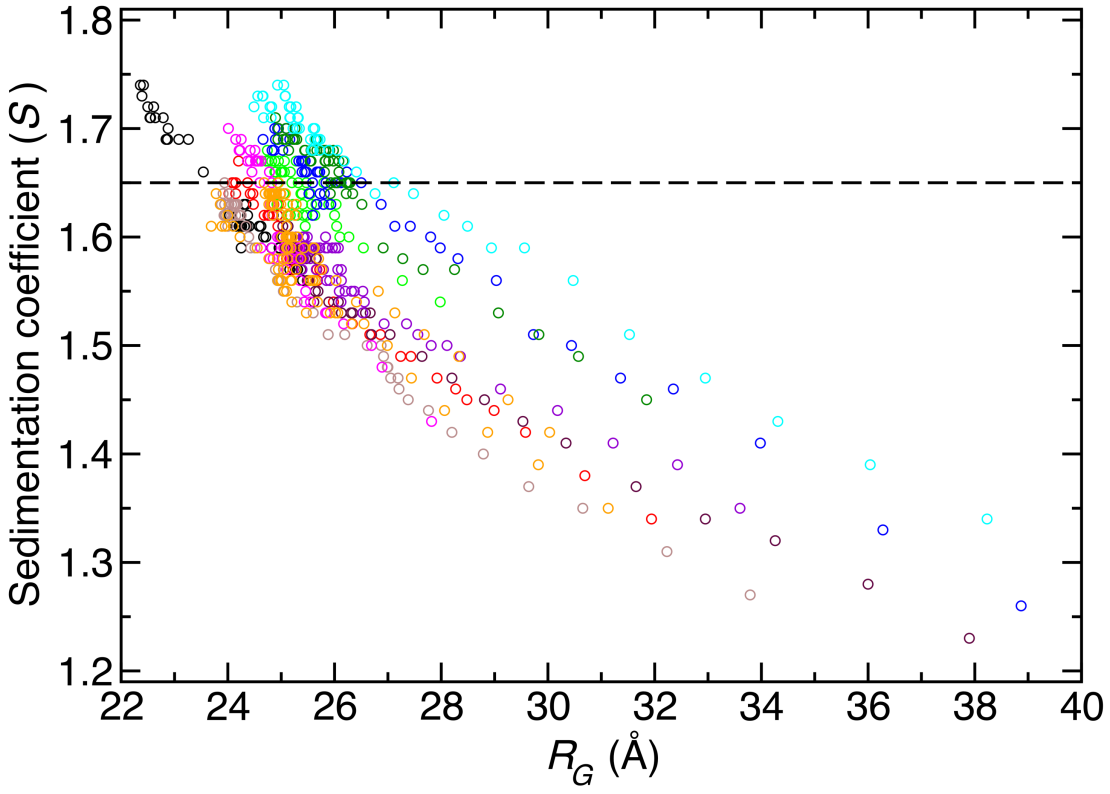
Figure 3.15 - Flowchart and Examples of apo uOmpA$_{171}$ structural models



Twelve independent uOmpA$_{171}$ models (residues 22-192) were created using the following protocol and are shown in the panel on the right. An initial, extended OmpA polypeptide (ϕ=-78°, ψ=149°) was constructed where amino acid residues were converted to a coarse-grained model with single pseudo-atom side chains. Torsion angles were altered using a Monte Carlo approach to obtain a randomly folded, but relatively expanded model (green and gold spheres). For the Monte Carlo procedure, new phi/psi values for a randomly chosen residue were attempted. The structure was filtered for atomic overlap, and the Metropolis criterion was applied with a scoring function that included residue-specific Ramachandran propensities (185) and backbone-backbone hydrogen bonding (186). After 2000 moves the structures were considered sufficiently folded for further collapse with molecular dynamics *in vacuo*. The initially folded course-grained model was converted to an all atom model (red, blue, white spheres) and further

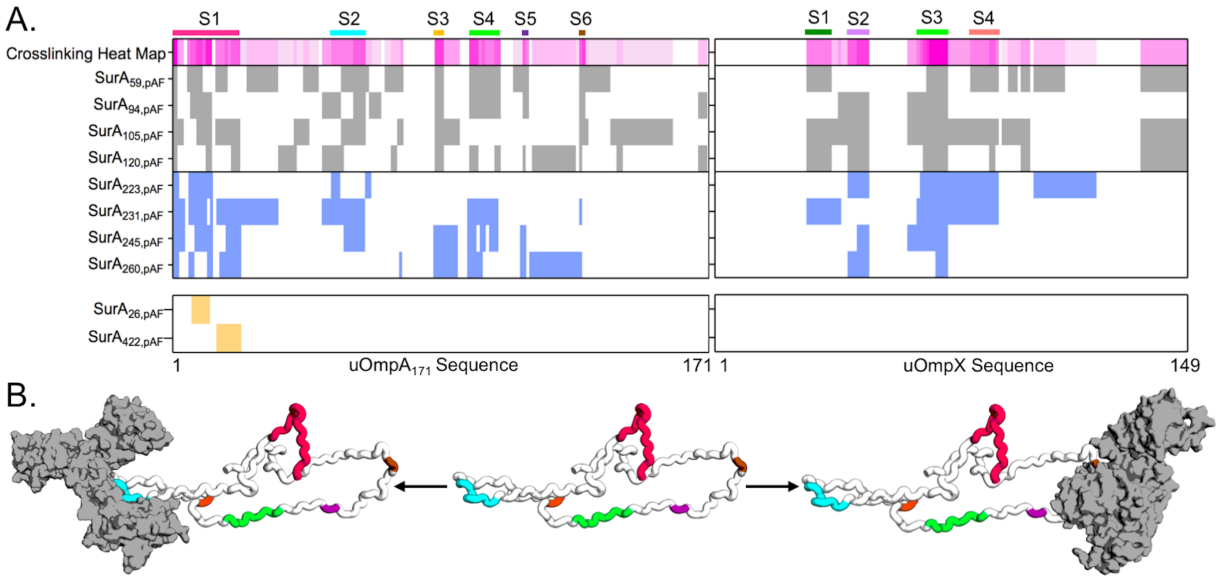
collapsed using molecular dynamics simulations in generalized Born implicit solvent with the collective variables module in NAMD (grey molecular surface). For the molecular dynamics collapse, 200 steps of energy minimization in the CHARMM22 force field were followed by 50,000 to 150,000 steps of MD with implicit solvent alpha cutoff=12.0 Å, [ion]=0.3M, non-bonded cutoff=14.0, switching starting at 13.0 and 2 fs time step. Langevin dynamics was used with a damping coefficient of 1 for temperature control (NVT). A collective variables radius of gyration biasing potential (lower wall=20.0 Å, upper wall=25Å) was used for final collapse using collective variables to drive molecular dynamics simulations (187). HullRad was used to calculate radius of gyration and sedimentation coefficient during the simulation and a structure was saved when R_G = 24.95±1.06 Å and s = 1.641± 0.078 (125).

Figure 3.16 - R_G vs. s-value for intrinsic uOmpA$_{171}$ models



Each uOmpA model created using course grained Monte Carlo folding was further collapsed using the collective variables module in NAMD under implicit solvent conditions. The relationship of calculated sedimentation coefficient and R_G during collapse is plotted with different colored circles for each of the 12 initial models. Structures with $s \approx 1.65$ were chosen as representative structures of the intrinsic uOmpA$_{171}$ conformation in solution without denaturant or chaperones present.

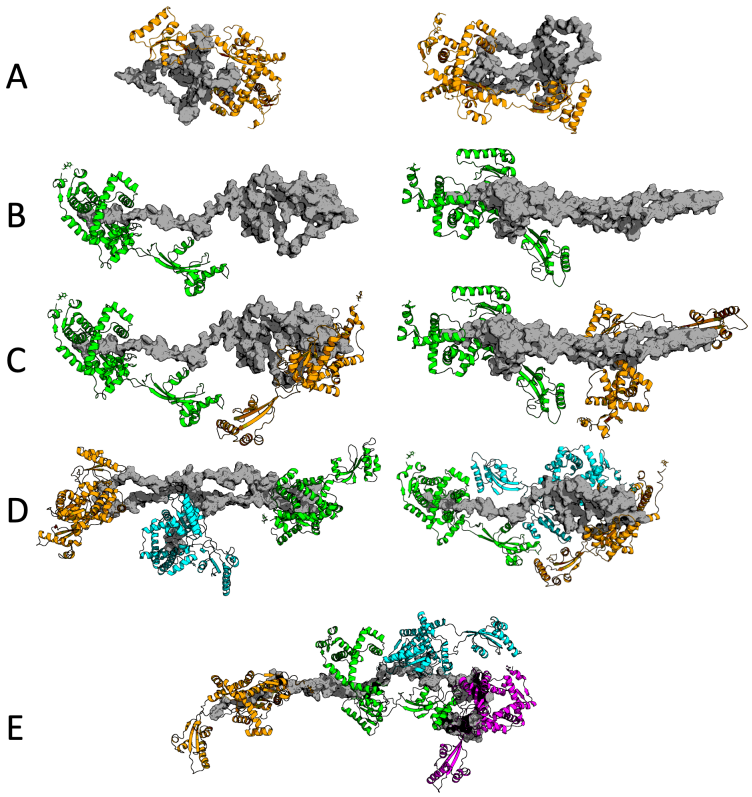
Figure 3.17 - Photo-crosslinking mass spectrometry (pXL-MS) identifies segments on client uOMPs that bind SurA.



(A) The crosslinking pattern for eight high-efficiency $SurA_{pAF}$ variants is shown to two uOMP clients ($uOmpA_{171}$ and uOmpX). Constructs that place *p*AF on the core domain are colored gray (top register) and constructs that place *p*AF on the P1 domain are colored blue (second register). The crosslinking heat map depicts the frequency a given residue on a client uOMP crosslinks to *p*AF in eight separate crosslinking experiments (darker magenta indicates residue is crosslinked more often). Binding segments are demarcated with a colored bar above the heat map and a label (S1-6 or S1-4). The bottom register shows results from two negative control studies using $SurA_{26,pAF}$ and $SurA_{422,pAF}$ (see main text for explanation). Only one $uOmpA_{171}$ crosslinked peptide was found for each of these constructs, while uOmpX did not crosslink at all. **(B)** An expanded $uOmpA_{171}$ model with hydrodynamic properties matching the contrast-matched SANS experiment is shown as a cartoon with the SurA-binding segments colored as in **(A)**. Two

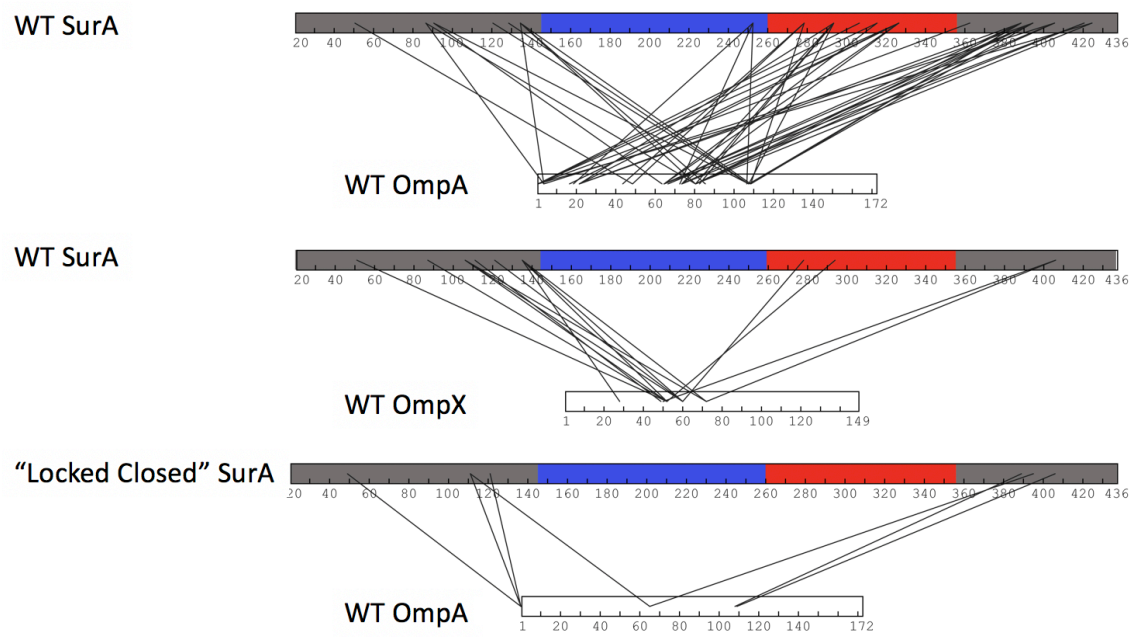
different segments (S2 and S6) are shown bound to SurA (shown in gray with a surface representation), suggesting that more than one copy of SurA could bind a single copy of $uOmpA_{171}$ with minimal steric clash.

Figure 3.18 - Example structural models of SurA-uOmpA$_{171}$ complex



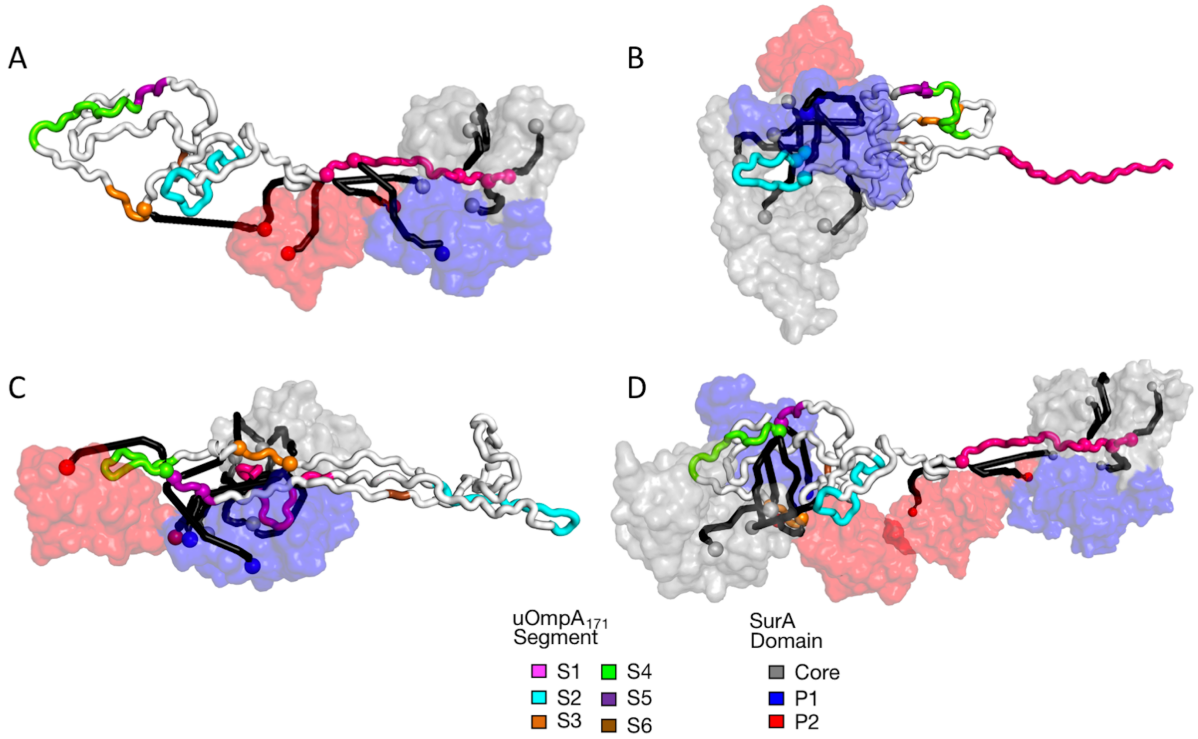
Representative snapshots of SurA-uOmpA$_{171}$ models used in the basis set for SANS analysis. **(A)** One SurA docked to non-expanded uOmpA$_{171}$ (showing 2 of 6 total); **(B)** one SurA docked to expanded uOmpA$_{171}$ (2 of 17); **(C)** two SurA docked to expanded uOmpA$_{171}$ (2 of 13); **(D)** three SurA docked to expanded uOmpA$_{171}$ (2 of 3); **(E)** four SurA docked to expanded uOmpA$_{171}$ (1 of 1).

Figure 3.19 - DSBU XL-MS Crosslinking shows client uOMPs bind in the SurA groove



DSBU crosslinks found between SurA and client uOMPs are shown (188). Sequences are shown as bars, with the SurA sequences colored based on the domain architecture outlined in Figure 3.1A. Lines between sequences represent a DSBU crosslink between residues, as determined by XL-MS (see Supplementary Data 3-5). The top two diagrams represent WT SurA crosslinking to each client uOMPs, $uOmpA_{171}$ and uOmpX. The bottom diagram contains the "Locked-Closed" SurA variant (P61C/A218C) in which a disulfide bond between the core and P1 domains inhibits the uOMP binding groove from forming. The difference in the total amount of crosslinks found comparing WT SurA and "Locked Closed" SurA mixed with $uOmpA_{171}$ shows that the formation of the groove is essential for efficient client uOMP binding.

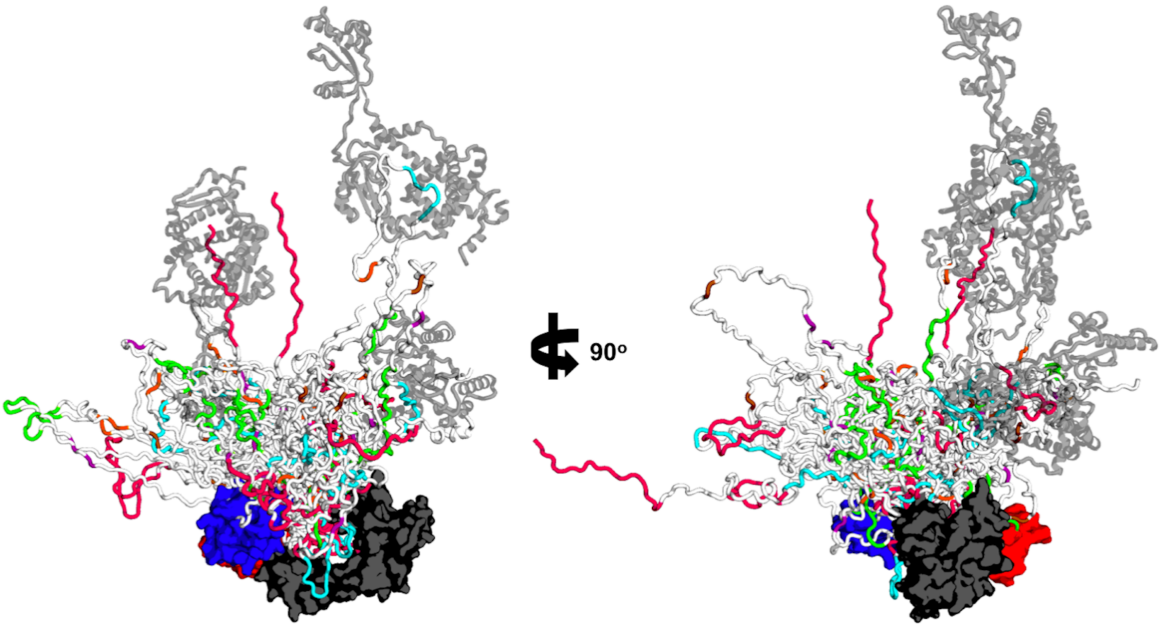
Figure 3.20 - Structural models of three SurA•uOmpA$_{171}$ binding modes validated by XL-MS and SANS.



SurA is represented as a surface (colored as in Figure 3.1A) and uOmpA$_{171}$ is represented as a tube (with binding segments demarcated by color as in Figure 3.17). DSBU crosslink sites are represented as spheres with crosslinks depicted as SASDs created by Jwalk. Crosslinks captured in each model are colored black (Figure 3.8 for details). (**A**) A representative structural model of the first binding mode (o1s016), supported by a cluster of ten crosslinks (pink) primarily between S1 and the core and P2 domains of SurA. (**B**) A representative structural model of the second binding mode (o1s010), supported by a cluster of nine crosslinks (blue) primarily between S2 and surrounding regions of uOmpA$_{171}$ and the core and P1 domains of SurA. (**C**) Representative structural model of the third binding mode (o1s009), supported by fifteen crosslinks between segment 1

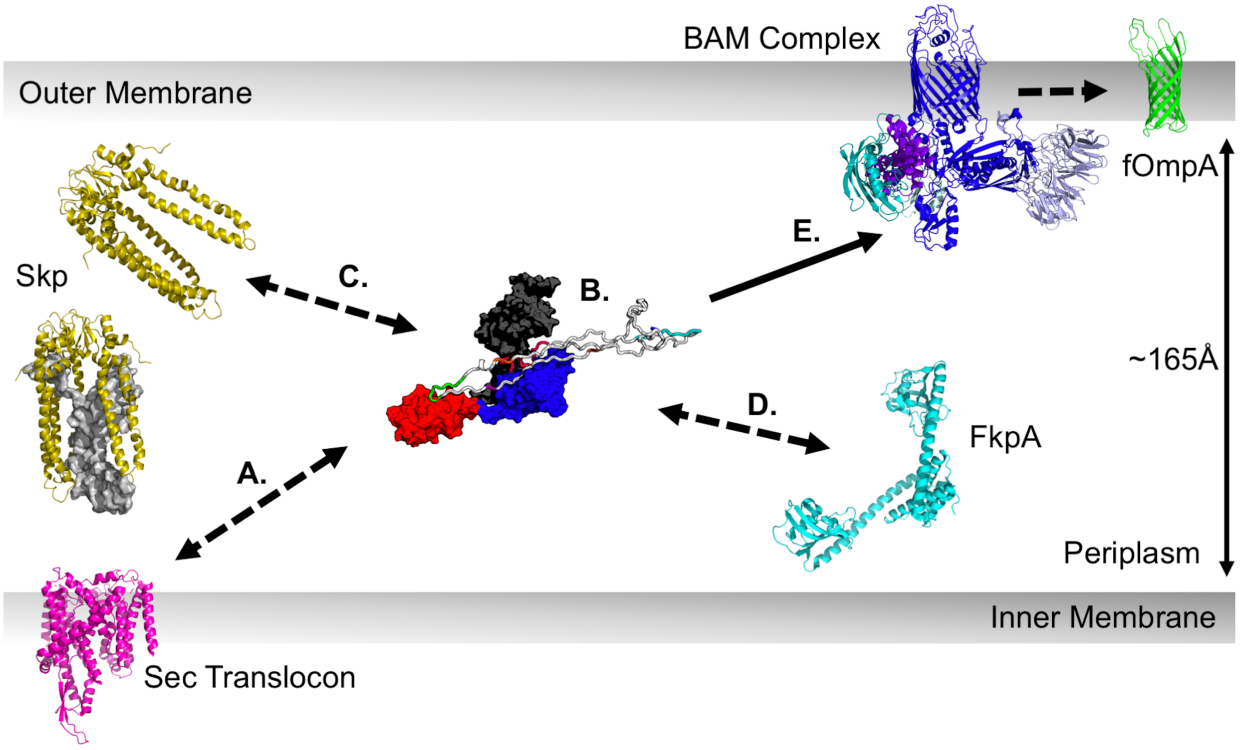
(pink) and segments 3, 4, and 5 (orange) primarily to the core domain of SurA. (**D**) A representative structure of two SurA protomers bound to $uOmpA_{171}$ (o2s006), wherein one copy of SurA binds segment 1, and another copy binds segments 3, 4, and 5.

Figure 3.21 - SurA•uOmpA$_{171}$ ensemble as defined by experimental restraints reveals uOmpA$_{171}$ conformational landscape.



The 21 SurA•uOmpA171 structural models that were part of triplets of linearly weighted models that fit the 0% D2O SANS data are overlaid aligned to SurA. SurA is in the open conformation and shown with a surface representation with domains colored as in Figure 3.1A. uOmpA171 models are shown with a cartoon representation and have SurA-binding segments, as defined by XL-MS, colored as in Figure 3.17. Higher order stoichiometries are found in the ensemble, with additional copies of SurA shown as transparent, gray cartoons. The diversity of uOmpA171 conformations that are shown in this ensemble highlight the conformational dynamics accessible to client uOMPs when bound to SurA.

Figure 3.22 - Implications of uOMP expansion in the periplasm.



uOMPs are post-translationally secreted through the Sec translocon (magenta), N-to-C terminally. The emerging uOMP N-terminus in the periplasm may be recognized by SurA. After complete translocation into the periplasm, one or more SurA protomers bind specific segments on uOMP clients, solubilizing the uOMP in highly expanded conformations roughly the width of the periplasm. The size of uOMPs bound to SurA is roughly double the size of uOMPs bound to Skp, one of the other chaperones in the network ($uOmpA_{171}$ bound to Skp shown as a gray surface representation in Skp closest to the translocon). The expanded uOMP may also be able to form heterocomplexes with other chaperones in the OMP biogenesis pathway: Skp (olive) and FkpA (cyan). The extended, unbound C-terminal region of the SurA-bound uOMP is also able to encounter the BAM complex, which recognizes the OMP β-signal and catalyzes uOMP folding into the outer membrane.

3.7 Tables

Table 3.1 - Parameters from Guinier Analysis of SANS Data

	Concentration (mg mL^{-1})	I(0) (cm^{-1})	R_G (Å)	R_G*q Range
$(SurA_{105}\text{-}uOmpA_{171})_{XL}$ (in 0% D_2O Hydrogenated SurA Deuterated $uOmpA_{171}$)	3.0	0.217 ± 0.002	44.0 ± 0.7	0.585 – 1.239
		0.215 ± 0.002	43.2 ± 0.7	0.621 – 1.265
		0.212 ± 0.003	42.6 ± 0.9	0.747 – 1.245
$(SurA_{105}\text{-}uOmpA_{171})_{XL}$ (in 30% D_2O Hydrogenated SurA Deuterated $uOmpA_{171}$)	3.0	0.054 ± 0.002	46 ± 3	0.615 – 1.254
		0.053 ± 0.002	45 ± 3	0.642 – 1.213
		0.053 ± 0.002	45 ± 3	0.696 – 1.271

Summary of the fitting parameters derived from Guinier fitting using a range of qR_G values for analysis. For $I(0)$ and R_G values, errors indicate the standard deviations from fitting. Rows shown in grey are included in Table 3.2.

Table 3.2 - Parameters from P(r) Analysis of SANS Data

	Concentration (mg mL^{-1})	D_{max} (Å)	I(0) (cm^{-1})	R_G (Å)	q Range (Å^{-1})
Prot-SurA$_{pAF}^{105}$ crosslinked to deut-uOmpA$_{171}$ (0% D_2O)	3.0	140	0.209 ± 0.001	42.0 ± 0.3	0.01436 – 0.1977
		150	0.212 ± 0.002	43.8 ± 0.5	0.01436 – 0.1977
		160	0.215 ± 0.002	45.1 ± 0.6	0.01436 – 0.1977
Prot-SurA$_{pAF}^{105}$ crosslinked to deut-uOmpA$_{171}$ (30% D_2O)	3.0	140	0.051 ± 0.001	44 ± 1	0.01436 – 0.1496
		150	0.052 ± 0.001	45 ± 1	0.01436 – 0.1496
		160	0.053 ± 0.001	46 ± 2	0.01436 – 0.1496
		170	0.054 ± 0.001	47 ± 2	0.01436 – 0.1496

Fitting parameters derived from generation of distance distribution functions, P(r) vs. r curves, for all SANS datasets using the specified D_{Max} values. For I(0) and R_G values, errors indicate the standard deviations from fitting. Rows shown in grey are included in Table 3.1.

Table 3.3 - Summary of all XL-MS and pXL-MS injections and FDR cut-offs.

	SurA Variant [a]	uOMP Variant [b]	XL[c]	Digest [d]	*n* [e]	N_{PSM} inter [f]	inter FDR cutoff[g]	N_{PSM} Intra[h]	intra FDR cutoff[i]
1	SurApaf105	OmpA171	paf	Trypsin	1	9	14.6	15	17
2	SurApaf105	OmpA171	paf	Glu-C	1	16	46.7	37	34.9
3	SurApaf105	OmpA171	paf	Glu-C	2	21	45.8	13	32.3
4	surApaf245	OmpA171	paf	Trypsin	1	7	15.3	2	14.5
5	surApaf245	OmpA171	paf	Trypsin	2	9	15	2	17
6	surApaf245	OmpA171	paf	Glu-C	1	23	26.2	1	34.4
7	surApaf245	OmpA171	paf	Glu-C	2	18	22.1	2	40.4
8	surApaf260	OmpA171	paf	Trypsin	1	17	17	1	14
9	surApaf260	OmpA171	paf	Trypsin	2	14	16.2	2	12
10	surApaf260	OmpA171	paf	Glu-C	1	7	14.5	1	19.8
11	surApaf260	OmpA171	paf	Glu-C	2	7	15.6	1	14.9
12	surApaf26	N/a	paf	Trypsin	1	N/a	N/a	0	11.2
13	surApaf26	N/a	paf	Trypsin	2	N/a	N/a	0	1.1
14	surApaf26	N/a	paf	Glu-C	1	N/a	N/a	0	11
15	surApaf26	N/a	paf	Glu-C	2	N/a	N/a	0	12.3
16	surApaf26	OmpA171	paf	Trypsin	1	0	N/a	0	N/a
17	surApaf26	OmpA171	paf	Trypsin	2	0	N/a	0	0.9
18	surApaf26	OmpA171	paf	Glu-C	1	1	14.9	0	12.6
19	surApaf26	OmpA171	paf	Glu-C	2	0	0.8	0	5.3
20	surApaf59	OmpA171	paf	Trypsin	1	49	23.1	25	28.8
21	surApaf59	OmpA171	paf	Trypsin	2	37	33.6	14	27.5
22	surApaf59	OmpA171	paf	Glu-C	1	44	43.6	6	41.4
23	surApaf59	OmpA171	paf	Glu-C	2	36	36.8	3	25.3
24	surApaf94	OmpA171	paf	Trypsin	1	24	23.7	8	18
25	surApaf94	OmpA171	paf	Trypsin	2	22	18.8	6	12.7
26	surApaf94	OmpA171	paf	Glu-C	1	17	13.8	2	16.6
27	surApaf94	OmpA171	paf	Glu-C	2	14	17.2	2	16.8
28	surApaf120	OmpA171	paf	Trypsin	1	24	18.7	14	40.8
29	surApaf120	OmpA171	paf	Trypsin	2	23	26.5	15	17.9
30	surApaf120	OmpA171	paf	Glu-C	1	55	23.5	8	23.4
31	surApaf120	OmpA171	paf	Glu-C	2	25	19.6	5	23.8
32	surApaf223	OmpA171	paf	Trypsin	1	2	12.7	17	16.9
33	surApaf223	OmpA171	paf	Trypsin	2	2	10	10	16.6
34	surApaf223	OmpA171	paf	Glu-C	1	3	20.1	29	20.3
35	surApaf223	OmpA171	paf	Glu-C	2	5	20.7	15	25.7
36	surApaf231	OmpA171	paf	Trypsin	1	6	14	2	11.7
37	surApaf231	OmpA171	paf	Glu-C	1	19	13.3	2	16.7
38	surApaf422	OmpA171	paf	Trypsin	1	0	15.4	1	18.4
39	surApaf422	OmpA171	paf	Trypsin	2	0	10.7	0	18.7
40	surApaf422	OmpA171	paf	Glu-C	1	0	12.9	0	22.1
41	surApaf422	OmpA171	paf	Glu-C	2	1	10.3	2	34.5
42	SurApaf105	OmpX	paf	Trypsin	1	11	10.9	0	11
43	SurApaf105	OmpX	paf	Glu-C	1	6	13.7	3	18.1
44	SurApaf245	OmpX	paf	Trypsin	1	3	4.5	0	2.4

45	SurApaf245	OmpX	paf	Glu-C	1	7	5.2	0	14.7
46	SurApaf59	OmpX	paf	Trypsin	1	10	8.3	3	10.5
47	SurApaf59	OmpX	paf	Glu-C	1	5	11.8	1	26.9
48	SurApaf260	OmpX	paf	Trypsin	1	3	8.1	0	0.1
49	SurApaf260	OmpX	paf	Glu-C	1	4	10.6	2	19.9
50	SurApaf26	OmpX	paf	Trypsin	1	0	1.7	0	13.5
51	SurApaf26	OmpX	paf	Glu-C	1	0	6	0	7.2
52	SurApaf94	OmpX	paf	Trypsin	1	5	11.5	1	13.5
53	SurApaf94	OmpX	paf	Glu-C	1	3	12.6	1	29.1
54	SurApaf120	OmpX	paf	Trypsin	1	16	15.2	18	22.8
55	SurApaf120	OmpX	paf	Glu-C	1	0	7.8	0	13
56	SurApaf223	OmpX	paf	Trypsin	1	6	4.5	15	15.1
57	SurApaf223	OmpX	paf	Glu-C	1	6	12.9	10	34.2
58	SurApaf231	OmpX	paf	Trypsin	1	7	10.3	6	12.7
59	SurApaf231	OmpX	paf	Glu-C	1	5	8.9	1	13.5
60	SurApaf422	OmpX	paf	Trypsin	1	0	0	0	9.9
61	SurApaf422	OmpX	paf	Trypsin	2	0	0	0	0.1
62	SurApaf422	OmpX	paf	Glu-C	1	0	9.2	0	9.3
63	SurApaf422	OmpX	paf	Glu-C	2	0	8.7	0	16.3
64	SurA WT	OmpA171	DSBU	Trypsin	1	10	0.2	58	2.2
65	SurA WT	OmpA171	DSBU	Trypsin	2	16	1.1	34	1.9
66	SurA WT	OmpA171	DSBU	Glu-C	1	24	0.9	25	1.2
67	SurA WT	OmpA171	DSBU	Glu-C	2	28	1.5	37	2.3
68	SurA WT	OmpX	DSBU	Trypsin	1	4	0.5	24	7.3
69	SurA WT	OmpX	DSBU	Trypsin	2	7	0.5	35	0.6
70	SurA WT	OmpX	DSBU	Glu-C	1	6	0.7	9	4.1
71	SurA WT	OmpX	DSBU	Glu-C	2	3	0.4	5	0
72	SurA PC2	OmpA171	DSBU	Trypsin	1	0	0	29	1.7
73	SurA PC2	OmpA171	DSBU	Trypsin	2	1	0	23	0.3
74	SurA PC2	OmpA171	DSBU	Glu-C	1	4	0	11	1.5
75	SurA PC2	OmpA171	DSBU	Glu-C	2	3	1.8	10	1.2

[a] Variant of SurA indicating the position of *p*AF. PC2 designates the locked-closed double-mutant P61C/A218C. [b] Variant of unfolded Outer Membrane Protein. [c] Crosslinker used (paf = para-azidophenylalanine; DSBU = disuccinimidyl dibutyric urea). [d] Designates whether trypsin digest was conducted alone, or trypsin and Glu-C digests were conducted in serial. [e] Designates which replicate (for conditions done in technical duplicate). [f] Number of interprotein peptide-spectrum matches in this injection. [g] The score above which interprotein peptide-spectrum matches achieve an FDR < 0.01. [h] Number of

intraprotein peptide-spectrum matches in this injection. [i] The score above which intraprotein peptide-spectrum matches achieve an FDR < 0.01.

Table 3.4 - Description of apo SurA Conformational Variants

Open	This conformation has both P1 and P2 domains open away from the core domain. It was built from 1M5Y, and the relative orientation of P1 to the core is the same as in 2PV3.
P1 closed	This conformation is based on the 1M5Y crystal structure. Missing loops and a C-terminal His tag were added using Modeller.(Fiser, Do, and Šali 2000)
P2 closed	The core-P1 orientation is based on the 2PV3 dimer structure, and the P2 is collapsed in the binding groove between core and P1.
Collapsed	The P1-core domain structures are the same as 1M5Y, with the P2 domain collapsed onto the core.

Table 3.5 - Hydrodynamic Description of SurA-uOmpA$_{171}$ Complex Models

Model ID	Total R_G	Total D_{MAX}	uOmpA$_{171}$ R_G	uOmpA$_{171}$ D_{MAX}
o1s001	35.41	112.79	26.23	83.62
o1s002	34.52	107.21	25.91	85.88
o1s003	38.68	137.14	33.87	111.55
o1s004	41.43	173.51	37.62	144.72
o1s005	33.29	109.93	21.61	65.54
o1s006	35.06	132.71	26.80	101.26
o1s007	38.97	167.94	39.84	155.36
o1s008	39.43	142.56	37.89	121.53
o1s009	41.86	169.14	40.48	148.80
o1s010	36.48	160.26	32.71	153.34
o1s011	35.89	155.36	39.84	155.36
o1s012	40.65	166.69	37.89	121.53
o1s013	41.51	175.68	40.48	148.80
o1s014	39.61	153.34	32.71	153.34
o1s015	38.07	170.90	34.95	150.85
o1s016	46.81	166.42	34.95	150.85
o1s017	37.36	118.28	27.98	84.00
o1s018	39.00	152.92	31.78	110.55
o1s019	39.87	147.24	31.78	110.55
o1s020	36.60	120.18	31.78	110.55
o1s021	48.48	194.40	41.99	147.32
o1s022	50.29	198.01	41.99	147.32
o1s023	39.83	147.32	41.99	147.32
o2s001	41.38	167.94	39.84	155.36
o2s002	51.59	186.91	37.89	121.53
o2s003	54.44	195.87	40.48	148.80
o2s004	52.05	162.63	33.11	152.57
o2s005	51.67	160.26	32.71	153.34
o2s006	58.59	187.61	34.95	150.85
o2s007	56.68	185.80	28.92	107.61
o2s008	42.48	170.60	35.10	150.65
o2s009	48.22	166.94	35.10	150.65
o2s010	41.30	152.92	31.78	110.55
o2s011	46.86	156.04	31.78	110.55
o2s012	78.34	246.64	41.99	147.32
o2s013	42.24	140.70	27.75	113.60
o3s001	54.20	185.34	35.10	150.65
o3s002	53.06	188.10	31.78	110.55
o3s003	69.95	246.64	41.99	147.32
o4s001	72.84	256.28	60.35	218.60

Highlighted models are found in the final ensemble of structures validated by XL-MS and 0% D_2O SANS experiments.

Table 3.6 Contrast Values for Experimental Components.

% D_2O	I(0) (cm^{-1})	C_{TOT} (10^{-3} g cm^{-3})	Contrast SurA (10^{10} cm^{-3})	Contrast 80% deut OmpA (10^{10} cm^{-3})
0	0.22	3.0	2.4	5.8
30	0.054	3.0	0.8	4.1

Table 3.7- SurA$_{pAF}$ Variant Crosslinking Efficiencies

	uOmpA (n=3-5)	uOmpX (n=3)	uOmpLA (n=2)
pAF Residue	Percent Crosslinked	Percent Crosslinked	Percent Crosslinked
D26	23.8 ± 6.3	23.7 ± 6.9	19.5 ± 3.4
Q47	0.6 ± 6.0	17.3 ± 11.4	6.7 ± 3.5
Q59	63.5 ± 7.0	62.1 ± 4.8	34.9 ± 7.5
E72	32.7 ± 5.6	28.7 ± 9.0	5.3 ± 6.4
K86	38.7 ± 9.8	24.7 ± 7.2	3.5 ± 6.9
E94	49.9 ± 9.5	41.6 ± 8.2	24.3 ± 6.4
K105	53.3 ± 13.1	51.8 ± 11.4	11.4 ± 11.1
Y120	50.9 ± 8.7	43.3 ± 4.1	13.7 ± 6.1
N126	17.3 ± 12.9	18.2 ± 13.3	5.5 ± 3.6
N144	38.0 ± 17.8	31.7 ± 8.4	9.9 ± 3.7
T151	29.3 ± 9.9	29.6 ± 2.7	2.3 ± 5.2
Q162	11.9 ± 6.2	10.6 ± 4.6	n/a
D190	20.9 ± 8.7	21.6 ± 5.3	n/a
R200	44.9 ± 12.0	33.9 ± 9.6	4.1 ± 6.1
H219	38.3 ± 8.3	33.9 ± 4.1	2.9 ± 4.0
Q223	50.4 ± 9.9	47.9 ± 3.2	11.3 ± 6.0
M231	53.3 ± 14.2	54.5 ± 4.4	14.9 ± 8.7
Q245	68.6 ± 6.6	68.9 ± 5.7	16.6 ± 7.3
K251	21.6 ± 4.7	18.2 ± 4.3	n/a
R260	67.7 ± 10.3	69.6 ± 5.1	n/a
K278	18.7 ± 12.1	22.9 ± 5.7	n/a
Q302	8.4 ± 8.2	3.3 ± 4.6	n/a
Q309	12.3 ± 11.7	7.2 ± 3.4	n/a
K326	41.1 ± 13.0	29.7 ± 4.9	n/a
W343	33.1 ± 5.9	19.1 ± 5.3	n/a
D350	25.9 ± 11.5	23.9 ± 9.9	n/a
R359	33.9 ± 13.2	33.6 ± 7.2	n/a
D382	5.1 ± 13.8	11.3 ± 5.0	2.3 ± 3.5
Y398	10.8 ± 13.5	11.4 ± 7.9	n/a
E408	31.3 ± 4.7	16.9 ± 5.1	n/a
M414	62.6 ± 8.5	31.1v± 15.1	n/a
Y422	8.2 ± 9.8	14.2 ± 3.6	1.39 ± 6.0

Crosslinking efficiencies for SurApAF variants crosslinked to three different outer membrane proteins. The errors reported for uOmpA$_{171}$ and uOmpX are standard deviation; the errors reported for the OmpLA values is the standard error of the mean. Values were considered not applicable if values were indistinguishable from controls

lacking SurA. Values were corrected for total intensity of $uOmpA_{171}$ lost due to UV alone ($uOmpA_{171}$ = 21%, uOmpX = 13%, uOmpLA = 22%).

Table 3.8 Members of each Triplet of Structures that Fit the 0% D_2O SANS Dataset.

Model 1	Model 2	Model 3
"P1 closed" SurA (42.4)	o1s005 (14.2)	o1s016 (43.4)
"P1 closed" SurA (45.6)	o1s013 (35.1)	o1s016 (19.3)
"collapsed" SurA (40.2)	o1s001 (13.4)	o1s016 (46.4)
"collapsed" SurA (33.3)	o1s002 (17.3)	o1s016 (49.3)
"collapsed" SurA (47.7)	o1s003 (17.4)	o1s021 (34.8)
"collapsed" SurA (45.6)	o1s004 (21.1)	o1s016 (33.3)
"collapsed" SurA (32.2)	o1s006 (19.2)	o1s016 (48.6)
"collapsed" SurA (39.0)	o1s009 (34.1)	o1s016 (26.9)
"collapsed" SurA (46.7)	o1s009 (28.0)	o1s021 (25.4)
"collapsed" SurA (43.6)	o1s009 (52.1)	o2s012 (04.3)
"collapsed" SurA (33.3)	o1s010 (20.1)	o1s016 (46.6)
"collapsed" SurA (35.6)	o1s012 (30.8)	o1s016 (33.5)
"collapsed" SurA (26.2)	o1s013 (52.3)	o1s016 (21.5)
"collapsed" SurA (34.5)	o1s013 (43.7)	o1s021 (21.8)
"collapsed" SurA (43.4)	o1s013 (32.7)	o1s022 (23.9)
"collapsed" SurA (33.9)	o1s013 (58.8)	o2s006 (07.3)
"collapsed" SurA (25.6)	o1s013 (71.0)	o2s012 (02.7)
"collapsed" SurA (36.8)	o1s013 (61.7)	o4s001 (01.5)
"collapsed" SurA (34.5)	o1s015 (21.8)	o1s016 (43.7)
"collapsed" SurA (44.5)	o1s016 (43.1)	o1s017 (12.3)
"collapsed" SurA (46.7)	o1s016 (38.5)	o1s019 (14.9)
"collapsed" SurA (43.4)	o1s016 (43.3)	o1s020 (13.3)
"P2 closed" SurA (49.8)	o1s001 (08.7)	o1s016 (41.5)
"P2 closed" SurA (45.6)	o1s002 (10.5)	o1s016 (43.9)
"P2 closed" SurA (32.2)	o1s005 (16.5)	o1s016 (51.3)
"P2 closed" SurA (44.5)	o1s006 (11.4)	o1s016 (44.0)
"P2 closed" SurA (48.8)	o1s007 (14.8)	o1s016 (36.4)
"P2 closed" SurA (49.8)	o1s009 (19.9)	o1s016 (30.3)
"P2 closed" SurA (45.6)	o1s010 (12.3)	o1s016 (42.1)
"P2 closed" SurA (47.7)	o1s012 (18.3)	o1s016 (34.0)
"P2 closed" SurA (37.9)	o1s013 (36.9)	o1s016 (25.2)
"P2 closed" SurA (47.7)	o1s013 (28.7)	o1s021 (23.5)
"P2 closed" SurA (32.9)	o1s013 (63.8)	o2s012 (03.3)
"P2 closed" SurA (45.6)	o1s015 (14.0)	o1s016 (40.4)

Each row of the table represents a combination of structures whose predicted scattering curves fit the experimental 0% SANS dataset (reduced chi-sq. < 1.05). Model 1 is always a conformation of SurA, which are detailed in Table 3.4. Models 2 and 3 are models of

the SurA-uOmpA$_{171}$ complex. Numbers in parentheses represent the weight percentage of each model in the triplet used to fit the experimental SANS data.

Table 3.9 Populations of Each Model in the SurA-uOmpA$_{171}$ Sparse Ensemble.

Model Name	Population (percent)
"P1 closed" SurA	2.51
"Collapsed" SurA	23.04
"P2 closed" SurA	15.09
o1s001	0.63
o1s002	0.80
o1s003	0.50
o1s004	0.60
o1s005	0.88
o1s006	0.88
o1s007	0.42
o1s009	3.83
o1s010	0.92
o1s012	1.40
o1s013	15.50
o1s015	1.02
o1s016	26.47
o1s017	0.35
o1s019	0.42
o1s020	0.38
o1s021	3.02
o1s022	0.68
o2s006	0..21
o2s012	0.31
o3s003	0.08
o4s001	0.04

Chapter 4 – Influence of Protein Scaffold on Side-Chain Transfer Free Energies

Published as: Marx, D.C., and Fleming, K.G., (2017) Biophysical Journal, 113(3):597-604

4.1 Introduction

Membrane protein folding involves the burial of amino acid side chains in lipid bilayers (189). The favorability of this process can be quantified as a transfer free energy (ΔG_{sc}°) from water to lipid for each side chain (190, 191). ΔG_{sc}° values are used to predict transmembrane domains from protein sequences and, in theory, changes to membrane protein stability due to mutations (75, 192). The accuracy of these predictions relies on the independence of measured side chain transfer free energies from the protein scaffold that they are located and experimental conditions in which they were determined. While the scaffold independence of side chain transfer free energies is always assumed, it has never been experimentally interrogated by measuring side chain transfer free energies in different scaffolds in the same experimental conditions.

Previously, we developed a host-guest system that allowed for side chain transfer free energies to be measured in the context of a natively folded protein in a lipid bilayer (97, 98). Using this system we measured side chain transfer free energies for the center of the bilayer using the transmembrane beta barrel E. coli outer membrane phospholipase A1 (OmpLA) as a scaffold protein (75). To assess the effects of protein scaffold on side chain transfer free energies we have extended our host-guest system to a new scaffold protein: the transmembrane beta barrel outer membrane acyl transferase PagP from E. coli. As a scaffold, PagP is structurally distinct from OmpLA in a variety of ways including the number of transmembrane β-stands, the tilt of the β-barrel with respect to the bilayer, and the amino acid composition of the transmembrane domain (26, 77, 193–195). In an effort to be consistent with the previously measured whole protein hydrophobicity scale,

we have chosen a host site on PagP that is predicted to be in a central and nonpolar region of the lipid bilayer. By measuring the at approximately the same depth in the bilayer and using the same experimental conditions as used previously, we can directly parse out the effects imparted on side chains by the scaffold protein.

We find that the PagP hydrophobicity scale correlates very well with the OmpLA hydrophobicity scale, especially for nonpolar residues. For polar residues we find energetic differences between the two proteins, with residues consistently more favorable in PagP than OmpLA. These discrepancies are rationalized as both local and global differences scaffold properties that alter ΔG_{sc}°. These findings highlight the intricacies of not only membrane protein folding, but also influence on individual side chain packing on the overall stability of membrane proteins.

4.2 Methods

PagP Expression, Purification, and Chemical Denaturation Titrations. The methods for protein cloning and purification, and experimental procedures are identical to previously published protocols (75, 76, 78). Briefly, WT PagP was previously cloned into a pET11A plasmid. PagP variants were cloned using the In-fusion HD cloning kit by Clonetech (Mountain View, CA) as described by the manufacturer, except V111C, which was cloned by Genewiz (South Plainfield, NJ). Plasmids were transformed into hms174(DE3) cells via electroporation and sequenced by Genewiz. Cells were grown in 500mL TB until OD600nm reached 0.8-1.0 and were induced with 1mM IPTG before growing for 6 hours. Cells were harvested by centrifugation at 5000 rpm for 15 minutes and stored at -20°C overnight.

Cell pellets were suspended in 25 mL lysis buffer (50mM Tris, 40mM EDTA, pH 8.0). Cells were lysed using Avestin emulsiflex C3 at 15,000 psi. Brij-L23 detergent was added at 0.1% before inclusion bodies (IB) were harvested via ultra centrifugation at 5500 rpm for 30 minutes, discarding the supernatant. IB pellets were suspended and pelleted two more times in wash buffer (10mM Tris, 1mM EDTA, pH 8.0) to wash out any contaminating, soluble macromolecules. Finally, IBs were suspended in 10 mL wash buffer and aliquoted into 10 USA Scientific microcentrifuge tubes. IBs were pelleted at 13.3 rpm for 5 min on a tabletop centrifuge and stored at -20°C after the supernatant was removed.

Before solubilization, IBs were suspended with 1mL titration buffer (100mM citrate, 2mM EDTA, pH 3.8), split into two microcentrifuge tubes, and pelleted at 13.3 rpm for 5

min. Inclusion bodies were solubilized in 8M guanidine hydrochloride titration buffer. Remaining insoluble contaminants were pelleted at 13.3 rpm for 20 min and the supernatant was filtered with a 0.22μm filter. Final protein concentration before folding proteins into LUVs was adjusted to ≤100μM.

DLPC lipids were purchased from Avanti Polar Lipids (Alabaster, AL) and aliquoted into acid washed glass vials at 25 mg/vial using Hamilton syringes. Lipids were dried with nitrogen gas before being put in a freeze dryer overnight and stored at -20°C under nitrogen. Lipids were solubilized to a concentration of 25 mg/mL in titration buffer. LUVs were formed via extrusion through a 0.1μm filter 21 times using an Avanti mini-extruder.

The procedure for folding PagP into LUVs was previously described and follows the procedure described for measuring OmpLA folding, except using the extinction coefficient of 82360 M-1 for PagP in 6M guanidine hydrochloride (GdnHCl) (76). Briefly, protein was diluted to 6 μM in the presence of SB3-14 detergent (Sigma) in 2.5M GdnHCl. Protein was then diluted to a protein to lipid ratio of 1:2000 in either 2.5M or 5M GdnHCl for unfolding and folding titrations, respectively. These samples were then incubated overnight at 37°C in a rotating incubator at 6 rpm. Samples were then finally diluted to 400nM protein concentration in microcentrifuge tubes at a range of GdnHCl concentrations (2M to 6M) and incubated for a minimum of 40 hours at 37°C in the rotating incubator.

Folding and unfolding of PagP was monitored via intrinsic tryptophan fluorescence. All fluorescence experiments were collected on an ISS PC1 photon counting fluorometer with excitation polarizer at 90° and 2.4mm slits, emission polarizer at 0° and 2.0mm slits,

and path length 1cm. For reversibility determination titrations, wavelength scans were collected (λ_{ex} = 295nm, λ_{em} = 280-400nm) twice for each sample. Full wavelength scans were fit to a log-normal function, as previously described. To determine PagP variant stabilities, 100 readings of emission intensity were recorded at each concentration of GdnHCl (λ_{ex} = 295nm, λ_{em} = 330nm). Three independent titrations of each PagP variant were collected.

To determine $\Delta G^{\circ}_{w,l}$ for all PagP variants, we globally fit all titrations to a two-state linear extrapolation model using Igor Pro (Wavemetrics). We determined a singular m-value that best described all PagP variants (m = 4.97), which was slightly different than the previously determined m-value for WT PagP (76).

Molecular Dynamics Simulations of PagP Variants. We used all-atom molecular dynamics simulations (MD) to investigate the structural consequences of the V111P PagP variant. Systems were constructed using 1mm4.pdb as the starting structure of PagP (1MM4) aligned by OPM (102, 195). CHARMM-GUI was used to build the systems and generate equilibration and production files (196–200). Two WT PagP and two V11P systems were generated with homogeneous DLPC bilayers and 0.2M NaCl. To approximate the pH of the experimental system (3.8), we protonated all histidine residues in PagP and left all Glu and Asp deprotonated. We chose to adopt this protocol because pH 3.8 is close to the model compound pKa values of Glu and Asp side chains, therefore accurately predicting the protonation states of these residues is challenging. Additionally, Glu and Asp residues in PagP are confined almost exclusively to soluble

regions of the protein, and should not affect the barrel conformation near the site of mutation.

Simulations were equilibrated using the CHARMM-GUI equilibration protocol. Trajectory 1 for both WT and V111P PagP were initially run at 30°C for 200ns. The temperature in the two systems was then slowly increased to 37°C to match the experimental temperature. These systems were run for 110ns more, of which the final 100ns were used in analysis. Trajectory 2 for both WT and V111P PagP were run using NAMD at 37°C for the entire simulation, again using the final 100 ns for analysis. Figure 4.1 shows that all trajectories equilibrated well before the final 100 ns that are used for analysis of the protein. All simulations were run on the Maryland Advanced Research Computing Center super computer. Analysis of the MD simulations was performed using a combination of VMD plug-ins and homemade scripts. To determine the effect of the V111P variant we measured distances between proposed backbone hydrogen bonding atoms (carbonyl-carbon and nitrogen) in residues 111 and 85 for the last 100 ns.

4.3 Results and Discussion

PagP Hydrophobicity Scale at site 111. We utilized a previously developed host-guest system to measure ΔG_{sc}° in which all residues are substituted into a chosen host site on PagP (75, 78). Figure 4.2 shows the location of the host site, residue 111 (valine in WT PagP), which is located at the end of the fifth strand of the transmembrane β-barrel. The nearest neighbor residues at this site are predominately hydrophobic (proline and alanine), with a tyrosine hydroxyl group as the lone source of polarity. As such, the host site on PagP is very chemically similar to the host site on OmpLA, yet it is composed of distinct residues. The z-distance from the bilayer center to the Cα for this site on PagP is predicted to be 5Å, which differs from the host site on OmpLA, which was calculated to 0.4Å (77, 194, 201). However, previous calculations of the depth-dependent atomic composition of symmetric DLPC bilayers found that the bilayer is essentially chemically identical within ± 5Å from the bilayer center (78). Additionally, side chains at site 111 extend towards the hydrophobic membrane center due to the tilt of the PagP barrel. Thus, the ΔG_{sc}° measured in PagP can be directly compared to those previously determined in OmpLA because the side chains are predicted to experience chemically identical regions of the bilayer. Because our host-guest system requires alanine to be the host residue, PagP V111A is the host protein that is used in all calculations. All other amino acids were individually substituted into this site as guest residues.

To extract thermodynamic quantities from folding titrations, PagP must fold spontaneously via a path independent mechanism. Wild-type PagP folding into 1,2-dilauroyl-sn-glycero-3-phosphocholine (DLPC) large unilamellar vesicles (LUVS) has

been previously shown to meet these requirements at pH 3.8 (76). Figure 4.3 shows we verified that all variants met these requirements using intrinsic tryptophan fluorescence, and Figure 4.4 shows representative high resolution, titration data demonstrating excellent fits by the two-state linear extrapolation equation, allowing for the stability in the absence of denaturant ($\Delta G^{\circ}_{w,l}$) to be calculated for each variant (117, 202). These experimentally observed values are reported in Table 4.1. By taking the difference in $\Delta G^{\circ}_{w,l}$ of the PagP host variant (Ala) and a PagP guest variant we can determine the transfer free energy ($\Delta\Delta G^{\circ}_{w,l}$) of the guest side chain into the center of the bilayer with respect to alanine (75, 78).

The resultant $\Delta\Delta G^{\circ}_{w,l}$ for all side chains are shown in Figure 4.5 and Table 4.1, where it can be observed that non-polar and aromatic side-chain substitutions are stabilizing to native PagP in bilayers. In contrast, ionizable and polar residues destabilized PagP structure. Given the highly hydrophobic composition of the DLPC bilayer around the host site, the burial of hydrophobic residues was expected to be favorable while hydrophilic residues were expected to be unfavorable.

Nonpolar Side Chain Burial at the Membrane Center is Scaffold Independent. For transfer free energies to be applied generally in predictive algorithms the transfer free energy for each residue should be an independent value. To remove the dependence of $\Delta\Delta G^{\circ}_{w,l}$ on the stability of the host protein (V111A), we calculated the nonpolar solvation parameter (σ_{NP}) for our host site in PagP (203, 204). Because σ_{NP} describes the linear

correlation between buried nonpolar surface area and transfer free energy, it can be used to calculate the theoretical ΔG_{sc}° for alanine.

We used the experimental data for the nonpolar side chains to derive σ_{NP} for PagP at site 111. Shown in Figure 4.6, this analysis revealed a robust, linear correlation between aliphatic side chain energetic perturbations and nonpolar accessible surface area in an extended G-X-G tripeptide ($R^2 = 0.91$). The slope reveals the magnitude of σ_{NP} to equal 25 cal^{-1} mol^{-1} Å^{-2} for transfer of nonpolar surface area from the membrane to water. This PagP σ_{NP} is almost identical to that observed in OmpLA (23 cal^{-1} mol^{-1} Å^{-2}) for a comparable depth in the bilayer (75). Therefore, we find the energy associated with nonpolar side chain burial in the center of the membrane is mostly independent of both local side chain packing and side chain orientation with respect to the membrane normal. Additionally, the PagP σ_{NP} is easily within the range of σ_{NP} measured in other contexts, such as hydrophobic solvents, α-helical membrane proteins, or statistical analyses of the interiors of soluble proteins, highlighting the robustness of our host-guest system (95, 203–205).

We used the nonpolar accessible surface area of the alanine side chain in a G-X-G peptide and this σ_{NP} to calculate the water-to-bilayer energy of alanine at our host site (ΔG_{sc}° = -1.73 kcal mol^{-1}). Adding this energy change to those of all side chains effectively shifts the set of measurements to a reference-free form. Figure 4.7 and Table 4.1 show these values (ΔG_{sc}°), which can be interpreted as side chain water-to-bilayer hydrophobicity values for all amino acid side chains.

Proline-Induced Secondary Structure Disruption is Energetically Equivalent in α-helices and β-barrels. By comparing the ΔG_{sc}° measured in PagP and OmpLA we can obtain a sense of the scaffold dependence of transfer free energy measurements of individual side chains. The biggest discrepancy between the two hydrophobicity scales is proline, which is the most destabilizing mutation in this investigation, but was found to be favorable in previous measurements ($\Delta\Delta G_{sc}^{\circ}$ = + 5.18 ± 0.14 kcal mol^{-1}) (75). This large difference in ΔG_{sc}° between the two studies highlights the magnitude to which factors other than solvation can influence the transfer free energy of a given side chain.

To investigate the molecular origins of the scaffold dependence of proline we performed MD simulations on both the PagP proline variant (V111P) and WT PagP in a DLPC bilayer. Because site 111 is at the end of a β-strand, we hypothesized that this location affects the energetic outcome through a mechanism independent of simple solvation. Indeed, proline is rarely found in secondary structure because of severe limitations on backbone torsion and the removal of a backbone hydrogen bond donor (206, 207). While the exposure of backbone polar atoms can be accommodated in soluble proteins through interactions with solvent, the center of the membrane lacks polar atoms to satisfy side backbone hydrogen-bonding requirements. In WT PagP residue 111 forms backbone hydrogen bonds with residue 85 in the neighboring strand (194, 195). To determine the effect of the V111P mutation on the structure of PagP we measured the distances between backbone nitrogen and carbonyl carbon atoms of residues 111 and 85 in both WT and V111P in multiple, independently seeded trajectories (208).

We found that the distances in V111P increase substantially compared to WT PagP, resulting in a break in secondary structure leading to local fraying of the ends and a localized unfolding event, which is not observed in the WT PagP trajectories (Figure 4.8). Thus, we find the exposure of multiple polar backbone atoms combined with other local unfolding events is the likely cause of the unfavorable ΔG_{sc}° for proline in PagP. We hypothesize that the host position in OmpLA can accommodate the torsion angle strain induced by proline because it is in the center of a strand rather than near the end of a strand, thus only breaking one transmembrane, backbone hydrogen bond. Additionally, the larger OmpLA has more water in the center of the transmembrane barrel, which may hydrogen bond with the exposed polar backbone atoms. Given estimates of a 4-5 kcal/mol penalty for breaking a hydrogen bond in the membrane, a simple interpretation of the $\Delta\Delta G_{sc}^{\circ}$ for proline could be that PagP V111P breaks an extra hydrogen bond compared to OmpLA A210P (191). Interestingly, the energetic cost for proline in PagP is almost identical to transfer free energies measured for proline insertion in the middle of transmembrane alpha helices, implying that the energetic cost of proline-induced structural deformations may be independent of secondary structure type (15, 207).

Comparison with OmpLA Hydrophobicity Scale Reveals Scaffold Dependence of ΔG_{sc}°. With the exclusion of proline, the PagP and OmpLA hydrophobicity scales are strongly linearly correlated (Figure 4.9). However, the slope of the correlation is not equal to 1 (slope = 0.91 ± 0.05), suggesting that ΔG_{sc}° for all side chains is not completely scaffold independent. In addition, the intercept is -0.83, indicating a context dependence of ΔG_{sc}°. We find that all side chains, except proline, have more favorable

ΔG_{sc}° when measured in the PagP scaffold than in OmpLA, though the magnitude of the difference in ΔG_{sc}° ($\Delta\Delta G_{sc}^{\circ}$) varies (Table 4.1). Because we find this trend applies for every side chain, general properties of the PagP structure must be contributing to the negative $\Delta\Delta G_{sc}^{\circ}$. Given the known structural differences, we hypothesize that the decreased strand number changes the overall density of side chain packing slightly and may increase side chain entropy or side chain contacts with lipid acyl tails resulting in ΔG_{sc}° that are more favorable in PagP.

For most nonpolar and aromatic residues we find $\Delta\Delta G_{sc}^{\circ}$ is less than 0.5 kcal mol^{-1}, suggesting that the differences between PagP and OmpLA are marginal for these residues (A, F, L, M, Y) and that ΔG_{sc}° is generally independent of the scaffold in which they were measured. This is unsurprising given the similar σ_{NP} determined for PagP and OmpLA. We find this scaffold independence also holds for glycine and serine. $\Delta\Delta G_{sc}^{\circ}$ for all other residues is greater than -0.5 kcal mol^{-1}, which indicates some prominent local or global difference between the two scaffold proteins that is altering the magnitude of ΔG_{sc}°.

We find a $\Delta\Delta G_{sc}^{\circ}$ of approximately -1 kcal mol^{-1} for beta-branched residues isoleucine, threonine, and valine (Table 4.1). We attribute this difference in ΔG_{sc}° to the evolved local side chain packing in each scaffold for the WT residues in each host site. Because the WT residue at the PagP host site is valine, we hypothesize that PagP has evolved favorable contacts between nearest-neighbor residues and branched beta-carbons at site 111. This contrasts with OmpLA, which has an alanine at the host site 210, and therefore would not have evolved favorable interactions for beta branched residues at that site.

The remaining residues that exhibit apparent scaffold dependence for ΔG_{sc}° all contain polar atoms, with $\Delta\Delta G_{sc}^{\circ}$ ranging from -0.6 to -2.0 kcal mol^{-1}. Mechanistically, solvation of polar residues in the membrane is accomplished in two ways: side chain "snorkeling" from the center of the bilayer to the interface, and the formation of lipid defects such as water dimples (77). While we initially thought that the placement of the host site on the tilted face of PagP would decrease the ability of polar side chains to snorkel, the more favorable ΔG_{sc}° in PagP compared to OmpLA seems to disprove this hypothesis. One way to increase the accessibility of side chains to the interface would be if the tilt of the PagP barrel were dynamic. Previous molecular dynamics studies of beta barrels have shown that the transmembrane domains can adopt a range of tilt angles with respect to the bilayer normal (77). PagP is known to be a "dynamic barrel" that can adopt multiple states and thus may be able to access a range of conformations with respect to the bilayer in thermal equilibrium (195, 209). The addition of a polar residue at site 111 may change the average tilt angle to one that allows for more efficient side chain snorkeling to the interface. Additionally, this dynamic behavior of PagP could also serve to decrease the energy of creating lipid defects and allow for water dimple formation easier than OmpLA. Thus, we contribute the majority of the scaffold dependence of ΔG_{sc}° to the general properties of the scaffold protein that dictate the ability of a side chain to reach a favorable chemical environment.

4.4 Conclusions

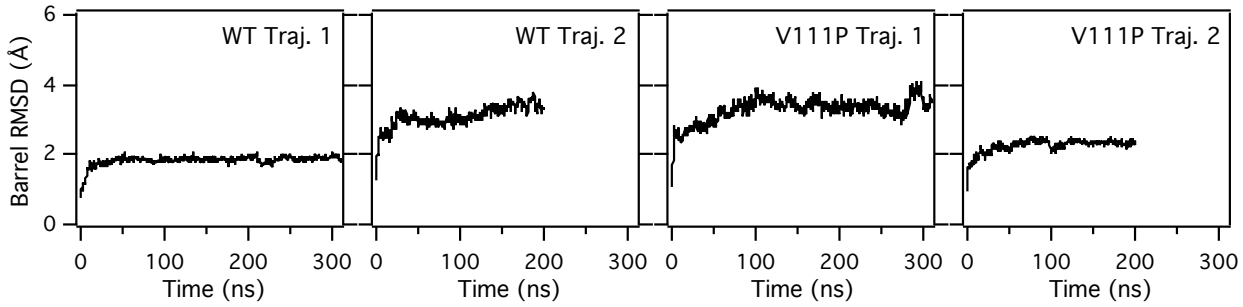
The energetics that govern membrane protein folding are linked to the burial of amino acid side chains in the hydrophobic lipid bilayer, which is theoretically independent of the protein sequence and structure. To investigate the validity of this assumption we measured ΔG_{sc}° in distinct scaffold proteins under the same experimental conditions for the first time. We find that while many residues exhibited differences in ΔG_{sc}° between measurements in PagP and OmpLA scaffolds, nonpolar residues, except valine and isoleucine, were found to be scaffold-independent. This finding indicates that the protein scaffold affects ΔG_{sc}° only when the side chain is placed in a non-compatible chemical environment and is forced to "snorkel" to a more favorable environment. The scaffold dependence therefore arises from the intrinsic properties of the scaffold protein and the bilayer that determine the energetic cost for side chain snorkeling. Additionally we found that the energy associated with proline-induced secondary structure disruption in PagP was very similar to those measured in a-helical membrane proteins. Together these findings highlight the general applicability of our whole protein hydrophobicity scales to nonpolar residues in the center of the bilayer and as well as the intricacies of side chain snorkeling and packing that complicate the process. We hope these results, which are applicable to all membrane proteins, will aid the production of accurate algorithms for predicting membrane protein stabilities from both sequence and structure.

4.5 Acknowledgements

This work was funded by NIH Grants R01 GM079440 and T32 GM008403. We thank Hector Figueroa and Anne Rice for assistance cloning the variants in this study, and Dr. Sarah McDonald and Dr. Patrick Fleming for experimental and computational guidance.

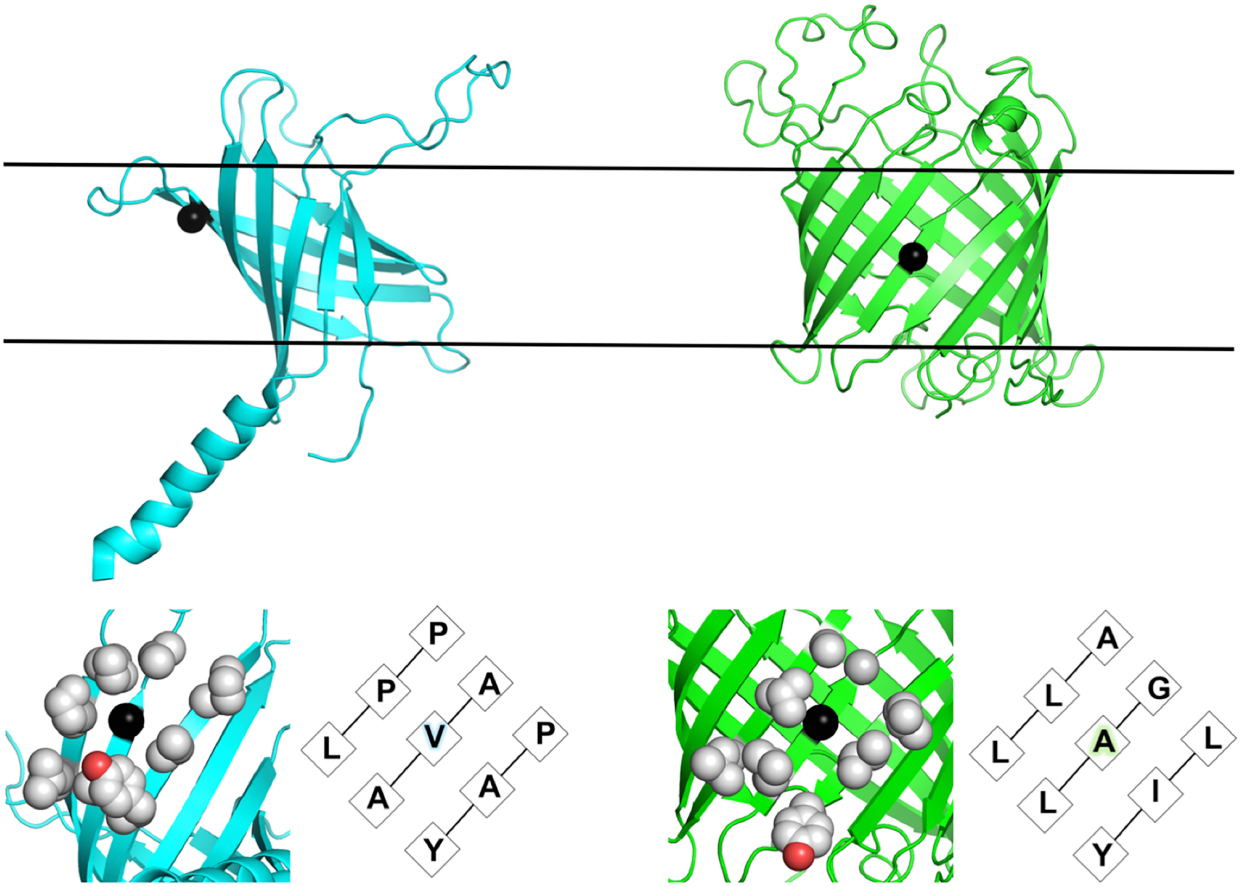
4.6 Figures

Figure 4.1 WT and V111P PagP molecular dynamics systems are equilibrated by 100 ns.



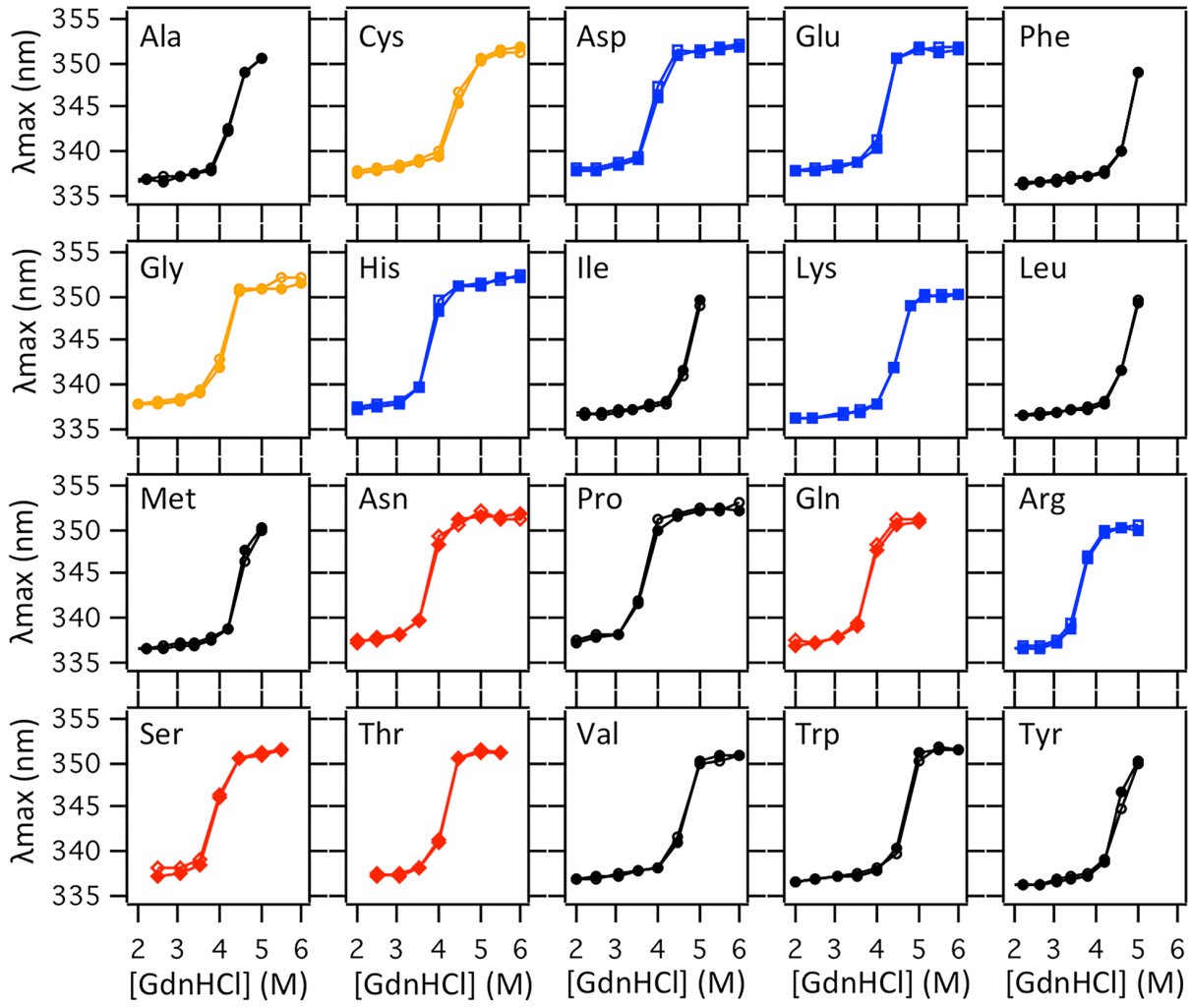
The RMSD (Å) compared to the starting structure of the backbone atoms are shown for all heavy atoms found in beta sheets. We find that all four trajectories equilibrated after 100 ns.

Figure 4.2 Comparison of structural characteristics of PagP and OmpLA.



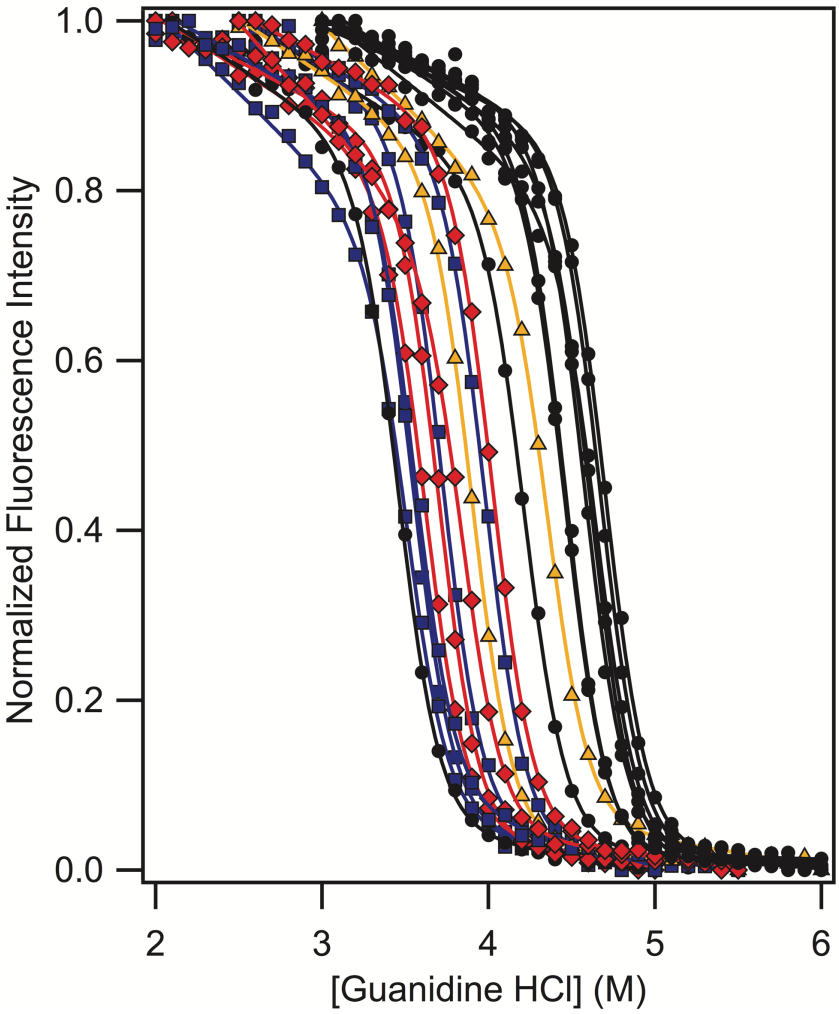
PagP (PDB: 1MM4), left, is colored in cyan, has 8 transmembrane beta strands and a periplasmic helix. OmpLA (PDB: 1QD5), right, colored green, has 12 transmembrane strands and is not tilted with respect to the bilayer (show as black lines). Sites at which transfer free energies were measured are shown as black spheres (PagP V111 and OmpLA A210), highlighting the different depths of the two positions. Shown below the structures are the nearest neighbor residues of the sites mutated in the two studies (shown in black spheres above). This figure was created using PyMOL (163).

Figure 4.3 All PagP variants show path independence in chemical denaturation titrations.



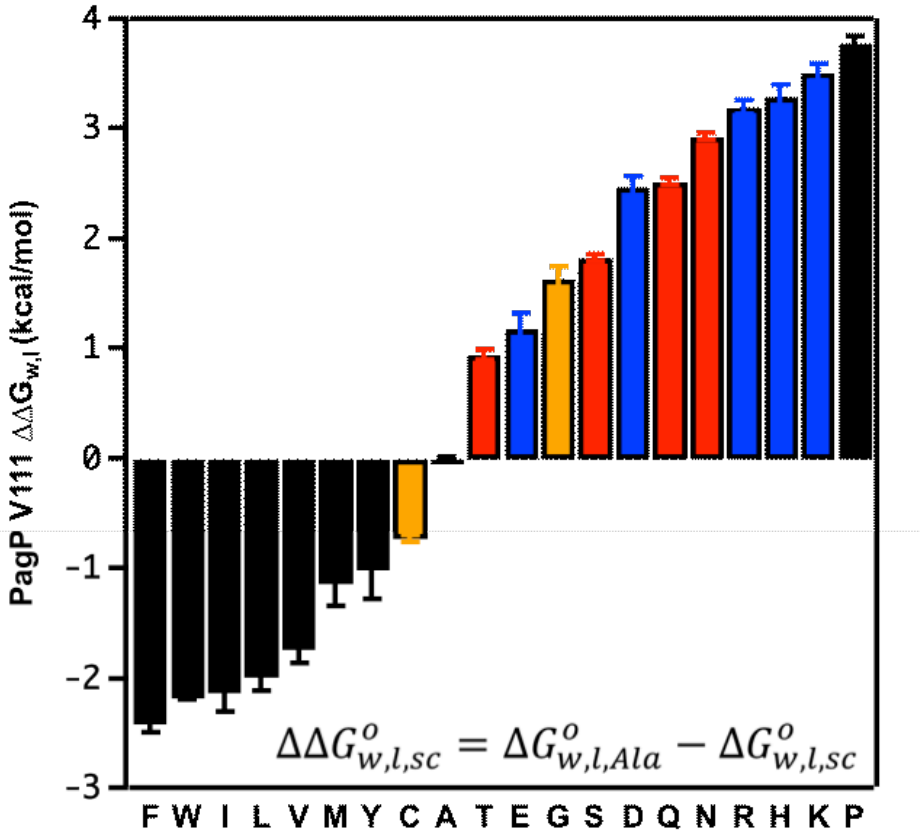
Unfolding titrations are shown by open symbols and overlaid upon folding titrations, which are shown by solid symbols. The variant is noted in each box, and the coloring scheme is identical to that in Fig. 4.4.

Figure 4.4 High-resolution chemical denaturation titrations of PagP variants.



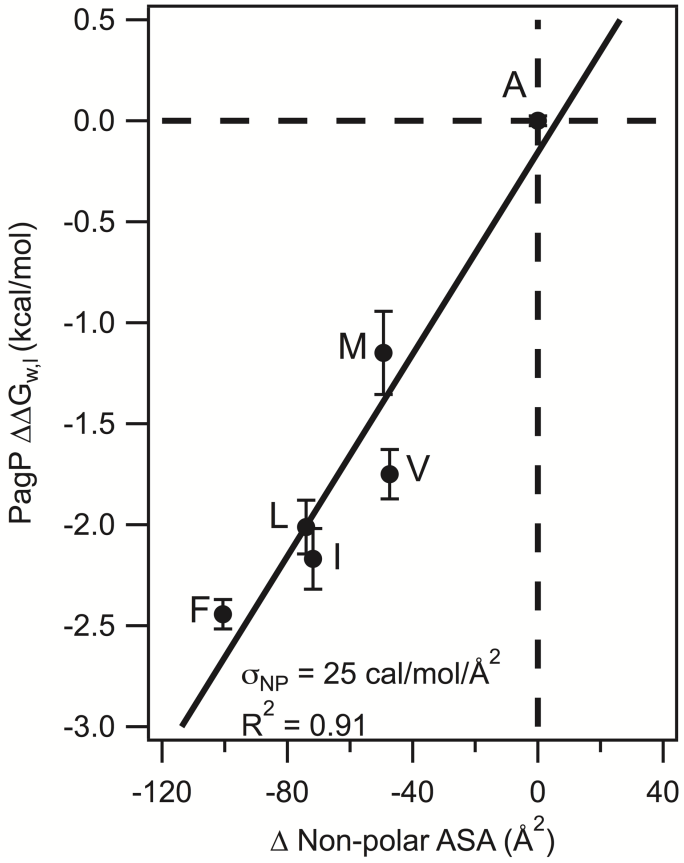
Representative chemical denaturation titration data for all 20 PagP variants at position V111 with ionizable residues colored blue, polar residues red, nonpolar residues black, and cysteine and glycine in gold. Normalized intrinsic tryptophan fluorescence intensity is plotted as a function of the concentration of guanidine HCl. Data were fit to a two-state folding model with the *m*-value held constant ($m = 4.97$).

Figure 4.5 PagP site 111 experimentally determined side chain energy values for all twenty amino acids.



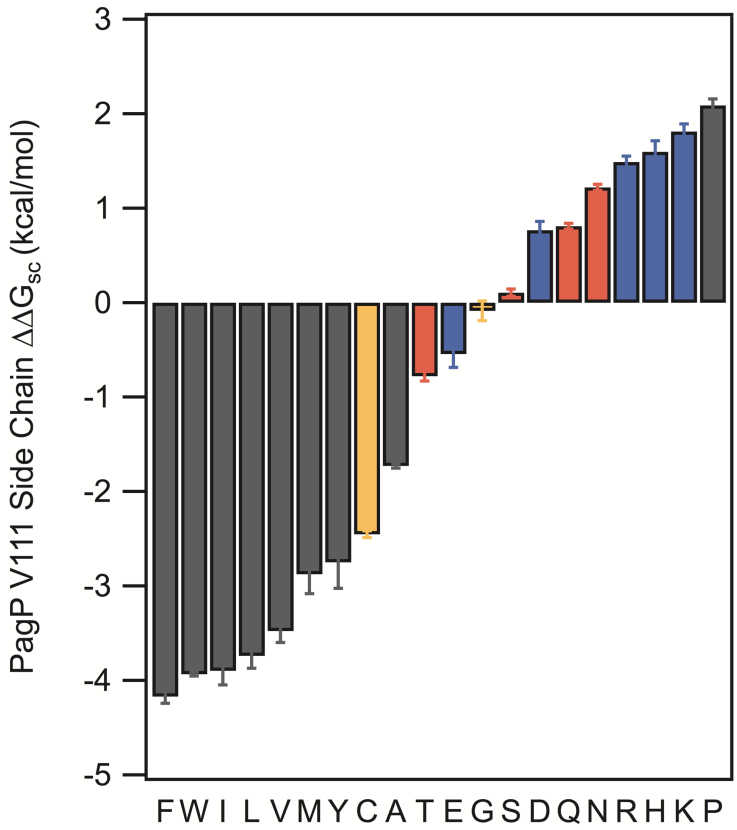
$\Delta\Delta G^{\circ}_{w,l}$ values were calculated by taking the difference in the stability of the V111A variant of PagP and V111X variant, where X is any amino acid (equation shown in figure). We find that most nonpolar residues are favorable at site 111 in PagP, except proline, which was the most unfavorable residue in the series. Polar residues were all unfavorable with respect to alanine, except cysteine. The coloring system follows the same convention as Figures 4.4.

Figure 4.6 Determination of the nonpolar solvation parameter at PagP site 111.



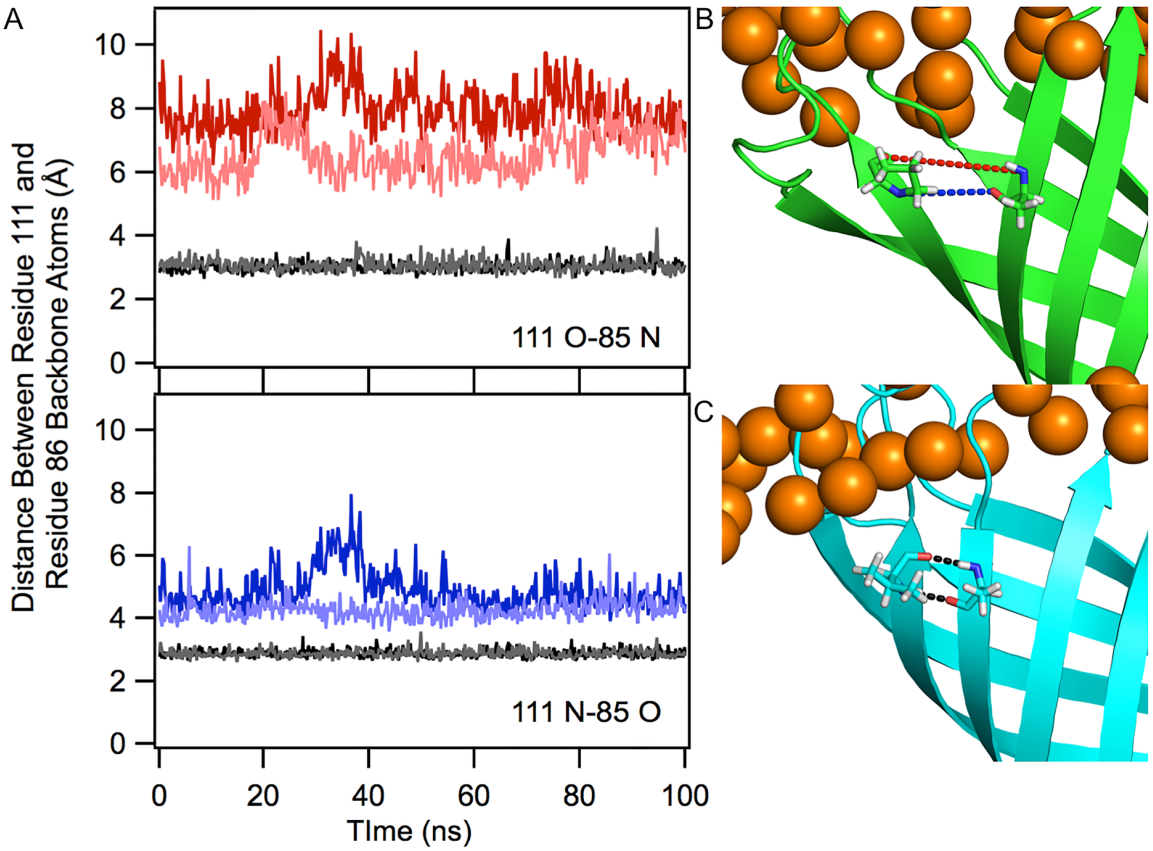
The $\Delta\Delta G^{\circ}_{w,l}$ values of nonpolar residues were plotted as a function of their nonpolar accessible surface area in a G-X-Gly tripeptide, where X is any amino acid. The correlation is well described ($R^2 = 0.91$), with intercept equal to 0.156 and slope equal to 25 cal mol^{-1} A^{-2}, which reflects the value of the σ_{NP} from bilayer to water.

Figure 4.7 Reference-free ΔG°_{sc} values for all side chains.



Side-chain transfer free energies for all 20 amino acids at site 111 after correcting for the transfer free energy of alanine. Error bars are standard deviations (n = 3 for all variants).

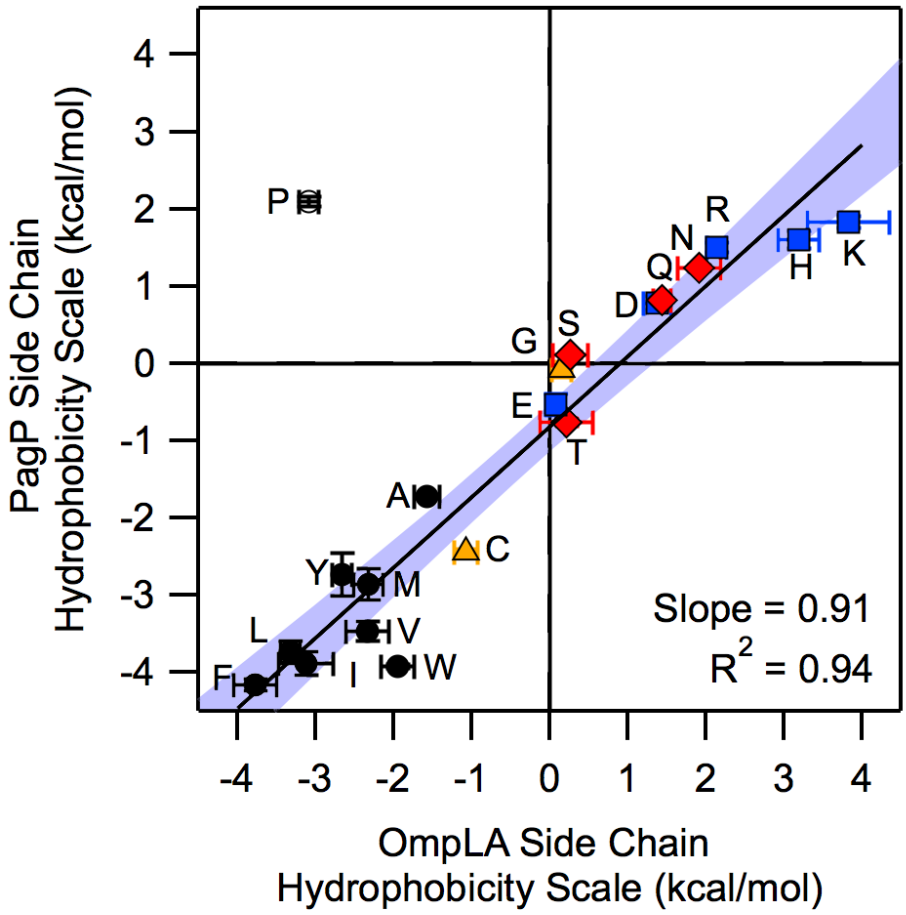
Figure 4.8 V111P induces local unfolding in PagP.



(A) To assess the effect of the V111P variant on PagP structure we measured the distance between the backbone hydrogen-bonding atoms be- tween residues 111 and 85. The top panel shows the distance between the carbonyl oxygen on residue 111 and the backbone nitrogen of residue 85, with the two WT trajectories colored gray and V111P trajectories colored red. The bottom panel shows the distance between the residue 111 backbone nitrogen and the residue 85 carbonyl oxygen, with WT PagP trajectories colored gray and V111P trajectories colored blue. The distances in V111P are too large for any contacts between the residues indicating a local unfolding event. (B) Snapshot from V111P trajectory showing the local unfolding of the β-sheet. The coloring scheme

from (A) is used to show the two different distances being measured. (C) Snapshot from the WT trajectory showing close contacts between residues 111 and 85. (B) and (C) were created using PyMoL, and DLPC phosphates are shown as orange spheres in both images.

Figure 4.9 PagP and OmpLA hydrophobicity scales are well correlated.



The plot shows the comparison of the two whole-protein water-to-bilayer hydrophobicity scales with PagP on the y axis and OmpLA on the x axis. Residues are colored with the same coloring scheme as in Figure 4.4. The 95% confidence intervals of the fit are shown in light blue. Proline was excluded from the linear fit and subsequent determination of confidence intervals.

4.7 Tables

Table 4.1 Values for Stabilities and Transfer Free Energies Variant

Variant	Experimentally Observed $\Delta G^{o}_{w,l}$ (kcal mol^{-1})	Difference between V111A and variant $\Delta\Delta G^{o}_{w,l}$(kcal mol^{-1})	Reference Free Side Chain Energies ΔG^{o}_{sc} (kcal mol^{-1})	Difference between PagP and OmpLA $\Delta\Delta G^{o}_{sc}$(kcal mol^{-1})
V111A	-20.96 ± 0.02	0.00 ± 0.02	-1.73 ± 0.02	-0.16 ± 0.02
V111C	-21.69 ± 0.03	-0.72 ± 0.04	-2.45 ± 0.04	-1.37 ± 0.15
V111D	-18.47 ± 0.09	2.49 ± 0.09	0.77 ± 0.09	-0.61 ± 0.20
V111E	-19.78 ± 0.14	1.18 ± 0.14	-0.54 ± 0.14	-0.61 ± 0.19
V111F	-23.41 ± 0.07	-2.44 ± 0.07	-4.17 ± 0.07	-0.40 ± 0.28
V111G	-19.32 ± 0.10	1.64 ± 0.11	-0.09 ± 0.11	-0.24 ± 0.16
V111H	-17.64 ± 0.12	3.32 ± 0.12	1.60 ± 0.12	-1.59 ± 0.28
V111I	-23.13 ± 0.15	-2.17 ± 0.15	-3.90 ± 0.15	-0.78 ± 0.38
V111K	-17.43 ± 0.08	3.54 ± 0.08	1.81 ± 0.08	-2.01 ± 0.53
V111L	-22.98 ± 0.13	-2.01 ± 0.13	-3.74 ± 0.13	-0.42 ± 0.19
V111M	-22.11 ± 0.20	-1.15 ± 0.21	-2.88 ± 0.21	-0.55 ± 0.28
V111N	-18.02 ± 0.03	2.95 ± 0.03	1.22 ± 0.03	-0.69 ± 0.28
V111P	-17.15 ± 0.07	3.82 ± 0.07	2.09 ± 0.07	5.18 ± 0.14
V111Q	-18.43 ± 0.03	2.54 ± 0.03	0.81 ± 0.03	-0.63 ± 0.11
V111R	-17.75 ± 0.06	3.22 ± 0.06	1.49 ± 0.06	-0.65 ± 0.14
V111S	-19.13 ± 0.03	1.83 ± 0.04	0.11 ± 0.04	-0.15 ± 0.22
V111T	-20.02 ± 0.05	0.95 ± 0.05	-0.78 ± 0.05	-0.99 ± 0.34
WT (V)	-22.71 ± 0.12	-1.75 ± 0.12	-3.48 ± 0.12	-1.14 ± 0.31
V111W	-23.17 ± 0.01	-2.21 ± 0.02	-3.93 ± 0.02	-1.98 ± 0.21
V111Y	-21.99 ± 0.28	-1.02 ± 0.28	-2.75 ± 0.28	-0.09 ± 0.30

Errors are standard deviations from n = 3 independent measurements.

[a]$\Delta G^{\circ}_{w,l}$ values correspond to the Gibbs free energy of folding of each variant from water (w) to the lipid bilayer (l) extrapolated to the absence of denaturant. A folding free energy *m*-value of -4.97 kcal mol^{-1} M^{-1} was determined as a global parameter by simultaneous fitting of all data in this study (Fig. 4.4).

[b]$\Delta\Delta G^{\circ}_{w,l}$ values correspond to water-to-lipid free-energy perturbations introduced by each mutation relative to the alanine variant (shown in Fig. 4.5).

[c]ΔG°_{sc} values correspond to side-chain energetic perturbations. These are reference-free values that have had the dependence of alanine removed by calculating the theoretical

ΔG_{sc}° of alanine using the nonpolar solvation parameter. The energies of all other side chains are adjusted by the new alanine energy (shown in Fig. 3.7).

[d]$\Delta\Delta G_{sc}^{\circ}$ values correspond to the difference between ΔG_{sc}° measured at site 111 in PagP and those measured at site 210 in OmpLA. Negative numbers indicate that ΔG_{sc}° values were more favorable (more negative) in PagP than in OmpLA. Positive numbers indicate that ΔG_{sc}° values were more favorable when measured in OmpLA. Errors were propagated from ΔG_{sc}°.

Chapter 5 –Local Bilayer Hydrophobicity Modulates Membrane Protein Stability

Published as: Marx, D.C., and Fleming, K.G., Local Bilayer Hydrophobicity Modulates Membrane Protein Stability. (2021) *Journal of the American Chemical Society*. doi: 10.1021/jacs.0c09412

5.1 Introduction

The cellular functions of a protein are regulated by the intrinsic thermodynamic stability of the three-dimensional structure adopted by its polypeptide chain (210–213). The stabilities of many water-soluble proteins have been experimentally measured *in vitro*, resulting in an understanding of the forces that regulate structure and function of this class of proteins (214). For membrane proteins, which make up approximately one third of the eukaryotic proteome (215), clarity on the balance of forces that determine their thermodynamic stabilities is much less mature (189). This lag in knowledge derives from the fact that membrane proteins are aggregation prone in aqueous solutions and must embed themselves into a hydrophobic environment such as detergent micelles or phospholipid bilayers to fold and function (16, 154, 216). The aggregation propensity renders traditional experimental approaches for measuring protein stabilities untenable for most membrane proteins (189).

The stabilities of membrane proteins are directly linked to the energetics of burying amino acid side chains into the phospholipid bilayer, also referred to as the side chain transfer free energy change (ΔG_{sc}°) (16, 73, 189, 191). Both the magnitude and sign of the water-to-bilayer ΔG_{sc}° for a side chain in a transmembrane domain (TMD) depend on the exact chemical compositions of the two end-point solvents (water and bilayer) (75, 78, 95, 96, 99, 157, 205, 217–219). There is the additional complexity in that a bilayer is not a simple end-point solvent. Biological membranes contain two chemically distinct regions: the largely desolvated hydrocarbon core and the chemically-heterogeneous interface (220, 221). Along the *z*-coordinate of the bilayer normal, each interface is an

approximately 15 Å wide region separating bulk water from the dehydrated center, and the transition between these two endpoints is characterized by a continuously changing water concentration (78, 221). This water gradient should result in a position-dependence on the free energies of transfer for side chains along the interface. Arguably, these membrane interfaces represent half of the bilayer volume, and a full understanding of membrane protein folding forces depends on the energetics imposed by this steeply changing polarity of this region.

The hydrophobic effect has empirically been related to the (water) solvent accessible surface area (ASA) of the side chain through an energy termed the nonpolar solvation parameter, σ_{NP} (75, 95, 157, 203, 222–224). This relationship quantifies the energy (cal mol^{-1}) gained per $Å^2$ of nonpolar surface area removed from water and buried in a nonpolar solvent. Across a wide range of model-protein systems and membrane mimics, the σ_{NP} value shows a remarkably precise value, ranging from -23 to -25 cal mol^{-1} $Å^{-2}$ (75, 95, 157). In the vast majority of studies, this singular σ_{NP} has traditionally been the basis for estimating side chain-bilayer partitioning as well as for identifying polypeptide sequence stretches that are likely to adopt transmembrane location. An exception to this is a single water-to-bilayer-interface study that found a σ_{NP} value for that reaction equal to -12 cal mol^{-1} $Å^{-2}$ (99).

One challenge with all of these studies is a precise mapping of the *z*-position in the membrane to which each σ_{NP} corresponds (99). Moreover, how to relate partitioning energies to a bilayer location is arguably impossible to know for those studies employing organic solvents as the bilayer mimic. Even with these drawbacks, usage of the nonpolar

solvation parameter and accompanying transfer free energy changes for nonpolar side chains have proven to be very powerful for identification of transmembrane segments from sequence.

Nevertheless, these energy values are still a somewhat crude tool for estimating the stabilities of these polypeptide regions at different locations in the chemically-heterogeneous bilayer. Here we take advantage of the naturally-occurring water gradient in the bilayer to ask how this changing water concentration affects the energetics of nonpolar side chains and accordingly, the nonpolar solvation parameter. We accomplish these measurements using a natively-folded protein that possesses a very favorable folding free energy and that is firmly anchored in its position within the experimental bilayer (77). Using this setup, we have resolved the bilayer *z*-position dependence of σ_{NP} using a combination of protein folding titrations, thermodynamic cycles and molecular dynamics simulations.

We discovered a linear correlation between σ_{NP} and the local water concentration across the bilayer normal. This relationship essentially quantifies the hydrophobic effect across the bilayer interface and enables prediction of the functional form for a depth-dependent solvation parameter, $\sigma_{NP}(z)$, and more generally, functions describing side-chain transfer free energies, $\Delta G^{\circ}_{sc}(z)$, across the bilayer. These energy values should improve the accuracy of membrane protein structure prediction and design algorithms. Additionally, we use our findings to better understand the thermodynamic basis for how the Sec translocon aids α-helical membrane protein folding *in vivo.*

5.2 Methods

OmpLA Variant Expression, Purification, Activity Assay and Folding Titrations. The OmpLA variants were engineered, expressed, and purified from inclusion bodies as previously described (75, 78). Each variant was verified to have enzymatic activity while folded into DLPC large unilamellar vesicles (LUV) using a previously developed colorimetric assay (75, 78). Folding titrations were also set up and performed as previously described in great detail (75, 78, 97, 98, 157). Briefly, unfolded OmpLA variants added to DLPC LUVs (diameter = 100nm) in 5M guanidine hydrochloride. After overnight incubation at 37°C, samples were diluted to varying guanidine concentrations ranging from 1-5M at 0.08M increments (51 total samples). Following at least a 40-hour incubation in a rotating incubator at 37°C, intrinsic tryptophan fluorescence was measured at 330 nm on a PC1 Fluorometer with cross-polarization (ISS, Champaign, IL, USA).

Fitting OmpLA variant $\Delta G^{\circ}_{w,l}$ and σ_{NP}. The resulting titration data were fit to a three-state linear extrapolation model to extract the stability ($\Delta G^{\circ}_{w,l}$) of each variant using previously determined *m*-values using the Igor Pro 8 software package (Wavemetrics, Portland, OR, USA).(75, 78) Host (alanine, Ala) variant $\Delta G^{\circ}_{w,l}$ were previously determined and used to calculate $\Delta\Delta G^{\circ}_{w,l}$ for each nonpolar guest variant.(78) The σ_{NP} for each site on OmpLA was determined by weighted linear regression using Igor Pro, with weights corresponding to the error (standard deviations) in $\Delta\Delta G^{\circ}_{w,l}$:

$$\Delta\Delta G^{\circ}_{w,l} = \sigma_{NP} * \Delta ASA + b \qquad \text{(Equation 5.1).}$$

where

$$\Delta ASA = ASA_{Guest} - ASA_{Ala}$$ (Equation 5.2).

and σ_{NP} and *b* are fitted parameters. The surface area for each nonpolar side chain was previously determined in the context of a Gly-X-Gly peptide.(75)

Molecular Dynamics Simulations of OmpLA variants. All-atom molecular dynamics simulations of each OmpLA variant using the CHARMM36 force field were performed to determine the position of each guest side chain in the experimental DLPC bilayer. The monomeric crystal structure of OmpLA (1qd5) was used as the starting structure, and the remaining components of the system and scripts were created using CHARMM-GUI.(196, 199, 200) A system was built for each OmpLA variant embedded in a DLPC bilayer and 0.2M NaCl at 37°C, resulting in 33 individual simulations. To approximate the experimental pH, some ionizable residues were protonated as described previously.(78) Systems were initially equilibrated using the protocol provided by CHARMM-GUI, followed by an additional 50ns of equilibration to allow for the system to fully relax (Figure 5.1). All simulations used NAMD and were run on the Maryland Advanced Research Computing Center super computer.(165)

VMD plug-ins and homemade scripts were used to analyze all trajectories.(208) To determine the position of the Cα for each variant relative to the phosphate plane, we first calculated the average *z*-position for the phosphate atoms in each leaflet of the bilayer. We then calculated the distance of the Cα for the variant side chain to the closest phosphate plane. These calculations were performed for each frame in the last 100ns of the trajectory, and the distances from the phosphate plane were binned into 0.1Å bins to create the histograms in Figure 5.2. Histograms were fit to Gaussian distributions using

Igor Pro software to determine average positions and standard deviations found in Table 5.1.

To relate the Cα positions of each side chain to the local concentration of water ([water]) in the bilayer we derived an empirical relationship between [water] and *z*-position in the bilayer (Figure 5.3). These data were derived from simulations of a neat DLPC bilayer and extracted using the Density Profile Tool in VMD.(225) We found that the data were well described by a sigmoidal function:

$$[water](z_{Phos}) = 1.0 * 10^{-6} + \left\{\frac{31.563}{1+exp^{\frac{3.7015-z}{-1.8604}}}\right\} \quad \text{(Equation 5.3).}$$

where *z* is the position relative to the phosphate plane of the bilayer.

The function can alternatively be adapted to any reference plane in the membrane by refitting $[water](z)$ with an adjusted z-position axis. For example, using the lipid carbonyl plane as the reference point (z = 0), the function changes to:

$$[water](z_{Carb}) = 1.0 * 10^{-6} + \left\{\frac{31.563}{1+exp^{\frac{-0.2985-z}{-1.8604}}}\right\} \quad \text{(Equation 5.4).}$$

Fitting and Derivation of $\sigma_{NP}([water])$, $[water](z)$, $\sigma_{NP}(z)$, and $\Delta G^{\circ}_{sc}(z)$. The relationship between σ_{NP} and [water] was determined by weighted linear regression (errors derived from error of fitting σ_{NP} using Equation 5.1) using Igor Pro, resulting in the following relationship:

$$\sigma_{NP}([water]) = 0.87 * [water] - 25.6 \quad \text{(Equation 5.5).}$$

The $[water](z)$ (Equation 5.3) function allows for $\sigma_{NP}(z)$ to be derived:

$$\sigma_{NP}(z) = 0.87 * [water](\mathrm{z}) - 25.6 \quad \text{(Equation 5.6).}$$

This function allows for $\sigma_{NP}(z)$ to be calculated at any *z*-position inside the phosphate plane of the bilayer. To derive $\Delta G_{sc}^{\circ}(z)$ in kcal mol^{-1} for any nonpolar side chain from $\sigma_{NP}(z)$ the following relationship is used:

$$\Delta G_{sc}^{o}(z) = \sigma_{NP}(z) * ASA/1000 \qquad \text{(Equation 5.7).}$$

where ASA is the solvent accessible surface area for that side chain.

Simulations and Analysis of Neat POPC Bilayer. A neat POPC bilayer (75 lipids per leaflet) was created using CHARMM-GUI with 0.2M NaCl at 37 ^{o}C. The CHARMM36m forcefield was used, and the simulation was equilibrated using CHARMM-GUI provided equilibration steps followed by 15 ns of additional equilibration time. The following 50 ns of simulation were used to measure the concentration of water molecules as a function of z-position in the bilayer and the z-position of the phosphate plane using the density profile plug in tool for VMD.(225) The relationship between z-position and [water] was fit to a sigmoid function as described above. The $\sigma_{NP}(z)$ for POPC was also determined as described above.

5.3 Results

Experimental determination of a z-dependent nonpolar solvation function. In this work we define a function, $\sigma_{NP}(z)$, that relates the value of the σ_{NP} to the *z*-position along the bilayer normal. This function is fundamentally derived from experimental protein folding titrations described below. Using a previously validated host-guest approach, we measured the thermodynamic stabilities of a series of host (Ala) and guest (nonpolar side chains isoleucine (Ile), methionine (Met), and valine (Val)) variants of the well-characterized membrane protein scaffold, E. *coli* outer membrane phospholipase A1 (OmpLA) (75, 78, 97, 98). Host/guest sites were chosen at varying positions across the bilayer to allow side chains to experience a range of local chemical compositions created by the amphipathic phospholipids in the membrane. In total, five lipid-facing sites located on well-defined transmembrane β-strands were used in this study (residues 120, 164, 210, 212, and 214) and are shown in Figure 5.4A. Figure 5.4B shows the water concentration gradient across the experimental DLPC bilayer and highlights the 60-fold change in water concentration accessed by these sites.

The thermodynamic stabilities ($\Delta G^{\circ}_{w,l}$) of all variants were measured using intrinsic tryptophan fluorescence-based chemical denaturation titrations (Figure 5.5) (75, 78, 97, 157). Figure 5.6 shows that each variant was enzymatically active, and Figure 5.7 shows that the folded and unfolded tryptophan fluorescence spectra for each variant overlay. Together these data indicate that single mutations do not affect the tertiary or quaternary structure of either the folded or unfolded states of OmpLA. We have previously shown that single site variants of OmpLA fold in a path-independent manner into large

unilamellar DLPC vesicles, allowing for equilibrium thermodynamic parameters to be extracted by fitting the titration data to a three-state linear extrapolation model ($\Delta G^{\circ}_{w,l}$ values listed in Table 5.2) (75, 78, 98).

The thermodynamic contributions of the guest side chain to folding were isolated by taking the difference in the $\Delta G^{\circ}_{w,l}$ values of the host and guest variants at a given site on OmpLA as shown in the thermodynamic cycle in Figure 5.8. We interpret the resulting $\Delta\Delta G^{\circ}_{w,l}$ at each site to represent the water-to-bilayer transfer free energy change for each guest side chain relative to alanine at a particular site (Table 5.2). We find that our experimentally measured $\Delta\Delta G^{\circ}_{w,l}$ are remarkably similar to analogous, computationally estimated $\Delta\Delta G^{\circ}_{w,l}$ that were previously published by Liang and coworkers, though both Ile and Val deviated significantly at site 212 (Figure 5.9) (226).

The nonpolar solvation energy is not constant along the bilayer interface. To determine the σ_{NP} for each site on OmpLA we used the $\Delta\Delta G^{\circ}_{w,l}$ values for each nonpolar side chain. This dataset included previously published $\Delta\Delta G^{\circ}_{w,l}$ for leucine and phenylalanine (75, 78). Figure 5.10 shows nonpolar $\Delta\Delta G^{\circ}_{w,l}$ values plotted as a function of the nonpolar surface area of each side chain relative to alanine (ΔASA) for each site. The σ_{NP} at each site on OmpLA is the slope of the linear fit of the correlation between nonpolar $\Delta\Delta G^{\circ}_{w,l}$ and ΔASA (Equations 5.1 and 5.2, fit parameters in Table 5.3). The σ_{NP} values vary nearly two-fold from site to site with a range that spans from -27.3 to -16.1 cal mol^{-1} Å^{-2}.

This experimental setup requires a reference host side chain whose energetic contributions must be removed to obtain a reference-free scale. As alanine is the amino acid side chain in the host protein, we obtain a reference-free scale by removing the intrinsic contributions of Ala to $\Delta\Delta G^{\circ}_{w,l}$ at each site using the site-specific σ_{NP} values. The reference-free ΔG°_{sc} for alanine at each host site was determined by multiplying the corresponding σ_{NP} by the ASA of the alanine side chain (75). The reference-free ΔG°_{sc} for the other nonpolar side chains were calculated by adding the ΔG°_{sc} for alanine to the $\Delta\Delta G^{\circ}_{w,l}$ at each site. The ΔG°_{sc} for each nonpolar side chain, including alanine, are listed in Table 5.4.

The nonpolar solvation parameter is linearly correlated with the amount of water in the bilayer. We next addressed the question of how the local chemical composition of the bilayer modulates the σ_{NP} by deriving the relationship between the σ_{NP} and *z*-position along the bilayer normal. This results in the function $\sigma_{NP}(z)$. To accomplish this, we calculated the *z*-position of the side chain Cα of each OmpLA variant from all-atom molecular dynamics simulations conducted in the experimental lipid bilayer. Previous work showed that introduction of single side chains has no discernable effect on the *z*-position or tilt of OmpLA in the bilayer (77). Figure 5.1 shows that each simulation was equilibrated after 50ns, and analyses were performed on the final 100 ns of each run. Instead of using the bilayer center as the reference point, $(z = 0)$, we calculated the position of each Cα relative to the average position of the phosphate plane of the bilayer.

We chose the phosphate plane as our reference point for z because it can be more easily applied to bilayers with different lipid compositions and acyl chain lengths. The

carbonyl plane is an alternative reference coordinate employed in the literature and could similarly be used. Figure 5.2 shows that variant Cα positions are generally well described by Gaussian distributions and that our sites capture the majority of the region between the bilayer center and phosphate plane. Table 5.1 lists the mean *z*-position and standard deviations for each variant Cα, which were averaged to calculate the *z*-position for the σ_{NP} at each site. We found that the bilayer *z*-position dependence of water concentration in our experimental DLPC bilayer was described by a sigmoidal function (Equation 5.3, Figure 5.3) (78). Using this correlation we calculates the corresponding local water concentration for each σ_{NP} (Table 5.5).

Figure 5.11A shows that the σ_{NP} is linearly correlated with the water concentration in the bilayer (Equation 5.5, R^2 = 0.84). The ~5,000-fold change in water concentration illustrates the large range of solvent conditions accessed in our experiments and necessitates presentation using a logarithmic abscissa. This novel functional relationship we derived from our measurements using a native protein and a phospholipid bilayer confirms our hypothesis that the water in a bilayer is a reporter on the hydrophobicity of the membrane and that the local hydrophobicity modulates membrane protein stabilities through the σ_{NP} directly.

It is also instructive to consider the σ_{NP} using bilayer *z*-coordinates. Figure 5.11B shows the bilayer z-position dependence of σ_{NP} in which the line of best fit in Figure 5.11A is transferred to units of distance from the phosphate plane of the membrane. Using the resulting function, $\sigma_{NP}(z)$ (Equation 5.6), the σ_{NP} can easily and accurately be determined at any position in the bilayer. $\sigma_{NP}(z)$ can also be applied to any membrane system with a

known [*water*] gradient along the bilayer normal. Figure 5.12 shows the [water](z) and $\sigma_{NP}(z)$ functions for a neat POPC bilayer, which is almost indistinguishable from DLPC indicating $\sigma_{NP}(z)$.

The $\sigma_{NP}(z)$ function plateaus to a value of -25.6 cal mol^{-1} Å^2 in the anhydrous core of the bilayer, which agrees with the σ_{NP} determined for this region of the bilayer in previous studies (75, 95, 157, 205). We find that the magnitude of the $\sigma_{NP}(z)$ decreases in the membrane interface to a predicted σ_{NP} equal to -1.4 cal mol^{-1} Å^{-2} at the phosphate plane, which is much less favorable than previous estimates of σ_{NP} for the interface (-12 cal mol^{-1} Å^{-2}) (99). Using $\sigma_{NP}(z)$, we calculate that this previously measured σ_{NP} for the water-to-interface transition corresponds to a z-position approximately 3.75 Å inside the phosphate plane. This *z*-position corresponds to the approximate position of the lipid carbonyl plane, which is a commonly used reference as it is approximately the boundary between hydrocarbon and interfacial regions of the bilayer.

Bilayer z-position dependence of reference-free ΔG°sc. The large change in the σ_{NP} across the bilayer interface significantly impacts the contribution of nonpolar side chains to the overall stability of membrane proteins in a position-dependent manner. This can be visualized by using the $\sigma_{NP}(z)$ to calculate a reference-free, nonpolar ΔG_{sc}° as a function of *z*-position in the bilayer (Equation 5.7). Figure 5.13 shows the $\Delta G_{sc}^{\circ}(z)$ for each nonpolar side chain (black lines) and the ΔG_{sc}° for each site on OmpLA determined above (points). Side chains located in the bilayer interface can be several kcal mol^{-1} less favorable than side chains in the dehydrated core of the bilayer. Additionally, our findings indicate that

interfacial side chains separated by a single turn of an alpha helix can have dramatically different contributions to the overall stability of a membrane protein.

The functional free energy dependence for the alanine side chain, $\Delta G^{\circ}_{Ala}(z)$, is particularly useful because the experimental setup necessitates that all side chain free energy perturbations be measured relative to Ala. Upon subtraction of the reference (Ala) energy, the *z*-position-dependence for any side chain can be place on a reference-free scale. Accordingly, the data herein allows conversion of previously measured $\Delta\Delta G^{\circ}_{w,l}$ values for arginine, tryptophan, and tyrosine to reference-free $\Delta G^{\circ}_{Ala}(z)$ scales (75, 78). Following the protocol above, a combination of new (Arg) and published (Tyr and Trp) all-atom molecular dynamics simulations were first used to determine the *z*-position of the Cα atom for each variant (78). This ΔG°_{Ala} value at a particular *z*-position was calculated using the alanine functional free energy dependence ($\Delta G^{\circ}_{Ala}(z)$), and this value was added to the corresponding experimental value for the guest side chain.

Figures 5.14, 5.15 and Table 5.6 shows these adjustments applied to the arginine, tryptophan, and tyrosine side chains (75, 78). As expected, these data show that there is a larger energetic penalty for the placement of the arginine Cα within the bilayer ~13 Å from the phosphate plane (+2.5 kcal mol^{-1}). This observation is in contrast to a small, favorable energetic contribution when the side chain is located in the bilayer interfacial region (-0.5 kcal mol^{-1}). The $\Delta G^{\circ}_{Arg}(z)$ is remarkably similar to the reported translocon-mediated profile from Arg insertion in a α-helical TMD (217). This similarity most like is derived from the fact that Arg can snorkel to the interface at any position in a TMD, which would change its thermodynamic “end state” to be the interface in all cases. The energetic

trends for the aromatic side chains $\Delta G^{\circ}_{Trp}(z)$ and $\Delta G^{\circ}_{Tyr}(z)$ reveal an opposite but also expected behavior: both side chains show an energetic preference for the bilayer interface as compared to the nonpolar central region of the bilayer.

The translocon energetically mimics the bilayer interface to facilitate TMD insertion into the membrane. With the reference-free, water-to-bilayer $\Delta G^{\circ}_{sc}(z)$ for nine amino acids determined as a function of *z*-position in the bilayer, we next asked whether our findings could be used to better understand membrane protein folding by the Sec translocon machinery. The translocon acts as a protein channel in the bilayer that allows α-helical TMDs to partition into the bilayer while protecting hydrophilic loops and turns from exposure to the hydrophobic membrane (16, 227). All-atom molecular dynamics simulations have identified that the chemical properties of water in the pore of the translocon mimic the properties of water in the interfacial region of the membrane (228).

Thus, a hypothesis for translocon function is that it mimics the thermodynamics of an interface-to-bilayer transition, not a water-to-bilayer transition (16). Figure 5.16A shows the relevant free energy reactions that can be written for these processes. The water-to-bilayer-center nonpolar ΔG°_{sc} values reported here are much greater in magnitude than the $\Delta G_{app\ (t,b)}$ for the translocon-to-bilayer transition. (15, 217). Because we have measured the water-to-bilayer ΔG°_{sc} across the bilayer normal, we can calculate the interface-to-center transition by taking the difference between ΔG°_{sc} at the center of the bilayer (A210) and the ΔG°_{sc} at the most interfacial site (Y214) (Figure 5.16A). Figure 5.16B shows that our interface-to-bilayer $\Delta\Delta G^{\circ}_{sc}$, or $\Delta G^{\circ}_{i,b}$, agree remarkably well with the

$\Delta G_{app\ (t,b)}$ (15). Additionally, the σ_{NP} for the interface-to-center transition ($\sigma_{NP,int-cen} = \sigma_{NP,214} - \sigma_{NP,210}$) is equivalent to the σNP calculated for translocon-mediated folding (7.7 and 6-10 cal mol^{-1} Å^2, respectively) (223). The remarkable energetic equivalence between the interface-to-center and translocon-mediated transfer free energies provides thermodynamic backing to the hypothesis that the translocon aids TMD folding by mimicking properties of the interface.

5.4 Discussion and Conclusion

The thermodynamic stability of a membrane protein regulates its structure, conformational landscape, and ability to perform its biological function (189, 191). A key determinant of this stability is the ΔG_{sc}° of each side chain from water into the membrane. Historically, experiments measuring side chain transfer from water to organic solvents such as cyclohexane or octanol were used to approximate ΔG_{sc}° (95, 96). While these experiments accurately describe the water-to-bilayer-center transition, they failed to capture thermodynamic properties resulting from the dramatic chemical heterogeneity of the interfacial region of the bilayer.

In this study we have measured the $\Delta\Delta G_{w,l}^{\circ}$ for Ile, Met, and Val as a function of z-position in the bilayer. Using these $\Delta\Delta G_{w,l}^{\circ}$, along with previously measured $\Delta\Delta G_{w,l}^{\circ}$ for Leu and Phe we were able to determine a novel functional form for the σ_{NP} across the bilayer normal, $\sigma_{NP}(z)$ (75, 78). We discovered a direct correlation between the σ_{NP} - and thus the ΔG_{sc}° for nonpolar side chains - with the local water concentration in the bilayer. The $\sigma_{NP}(z)$ is a convenient numerical expression for assessing the magnitude of the hydrophobic effect, as it relates nonpolar surface area to the energy gained from transferring the surface area from bulk water to a dehydrated environment.

Importantly, the dependence of nonpolar ΔG_{sc}° on local water concentration differs starkly from the water-dependence of backbone hydrogen bond (bbHB) formation in membrane proteins. Although it was long expected that bbHB energies would also be

strongly influenced by the steep water gradient intrinsic to bilayer interface, we recently observed bbHB energies to be constant throughout this region (79).

The experimental data here offer an opportunity to consider how the translocon insertion process relates to the underlying physical chemistry of the reaction. Our data provides evidence that translocon-mediated membrane protein folding energetically mimics the interface-to-bilayer transition for ΔG°_{sc}. We find that $\Delta G^{\circ}_{i,b}$ is approximately equal to $\Delta G_{app\ (t,b)}$ for eight of the nine side chains investigated, with Trp being the only outlier; we speculate that this one outlier is most likely due to differential aromatic interactions in the OmpLA and translocon experimental setups (15, 78). This finding provides thermodynamic support for a recent model for translocon function in which transmembrane α-helices sample both interfacial and translocon associated conformations before they insert into the membrane (16). Our findings indicate that these two states would be essentially energetically equivalent, indicating the energy barrier between these two conformations may be non-existent with switching between the two states determined by the kinetics for a given TMD sequence.

Our results should improve the speed and accuracy of membrane protein structure prediction and design algorithms (229–231). As membrane proteins account for over half of all therapeutic drug targets (232), having accurate force fields to describe membrane protein energetics is essential. This simple, linear function $\sigma_{NP}(z)$ could be adapted to implicit membrane models or could alternatively be applied explicitly in systems such that a local value of σ_{NP} to be calculated over water concentrations ranging from millimolar to

30 molar. Subsequently, better estimates of the energetic contribution of any lipid-facing, nonpolar side chain to the stability of membrane protein.

By relating the σ_{NP} to the chemical properties of the bilayer, the $\sigma_{NP}(z)$ can be customized to any bilayer lipid composition by calculating the water concentration gradient of that particular bilayer (illustrated in Figure 5.12 for a POPC bilayer). We find that the $\sigma_{NP}(z)$ for both POPC and DLPC bilayers are almost indistinguishable, indicating that lipid acyl tail length and saturation have little effect on the shape and magnitude of $\sigma_{NP}(z)$ across the bilayer interface. Further work is needed to assess the impact of headgroups and other macromolecules found in bilayers such as cholesterol on the shape of $\sigma_{NP}(z)$. In addition, one caveat of $\sigma_{NP}(z)$ is that it relies on additive forcefields, which may underestimate the local water concentration in the bilayer interface. As polarizable forcefields become more readily accessible the $\sigma_{NP}(z)$ could be refined to even more accurately quantify the energetics of nonpolar side chain burial in membranes.

The $\sigma_{NP}(z)$ and the nine $\Delta G_{sc}^{\circ}(z)$ functions derived here also could be applied to predicting the contribution of the binding energy of proteins to the bilayer interface ($\Delta G_{sc}^{\circ}(z)$ functions found in Table 5.7). By relating energy to *z*-position of side chains in the bilayer, binding energies can be tuned based on structural information derived from experiment or computation. The simplest application of the $\Delta G_{sc}^{\circ}(z)$ functions would apply to proteins that do not undergo large structural changes upon membrane interaction, as the binding energy would be dominated by lipid-side chain interactions. In cases of conformational changes, our values may find utility in providing baseline values for the partitioning of many membrane binding proteins (such as antimicrobial peptides) even

when the global binding energy contains additional contributions from other sources, such as the energy of the coil-to-helix transition that occurs upon binding (75). Given the complexity of membrane protein folding, even for apparently simple peptides, the safest application of our values is in understanding the effect of a point mutation on bilayer-protein interactions.

Additionally, our findings delineate how the different transfer free energies reported here should be applied to various biological functions. The critical parameter for consideration is the endpoints for the reaction at hand. The $\Delta G_{app\,(t,b)}$ or $\Delta G^{\circ}_{i,b}$ can be used to approximate the contribution of a side chain to the folding of alpha-helical TMDs (15, 217). On the other hand, water-to-bilayer ΔG°_{sc} are more applicable to modeling both changes in stability of different conformers of a membrane protein, the transition between interfacial and transmembrane conformations of antimicrobial peptides, and potentially unfolding processes, such as extraction of transmembrane regions from the bilayer by AAA-ATPases (233). Experimentally determining these thermodynamic parameters for membrane proteins is extremely challenging and accurate computational modeling will allow for greater understand of essential processes such as the gating of a channel or the cycling of a transporter (234–236). At the very least, we anticipate that $\Delta G^{\circ}_{sc}(z)$ can be used to better understand the effects of mutations that may impair membrane protein folding such as those occurring in cystic fibrosis and Charcot-Marie-Tooth disease (237).

In summary, we have determined the relationship between nonpolar side chain transfer free energies, solvent accessible surface area, and the local water concentration in the bilayer. Using this relationship, we calculate $\Delta G^{\circ}_{sc}(z)$ functions for nonpolar side

chains and remove the alanine dependence of previously measured $\Delta\Delta G^{\circ}_{w,l}$ for Arg, Trp, and Tyr. Our work shows that the translocon energetically mimics the bilayer interface for side chain transfer into the membrane. Together, these findings increase our understanding of the driving forces of protein stability in membranes by essentially quantifying the hydrophobic effect along the bilayer normal and should increase the accuracy of computational workflows to identify, design, and understand the energy landscapes of membrane proteins.

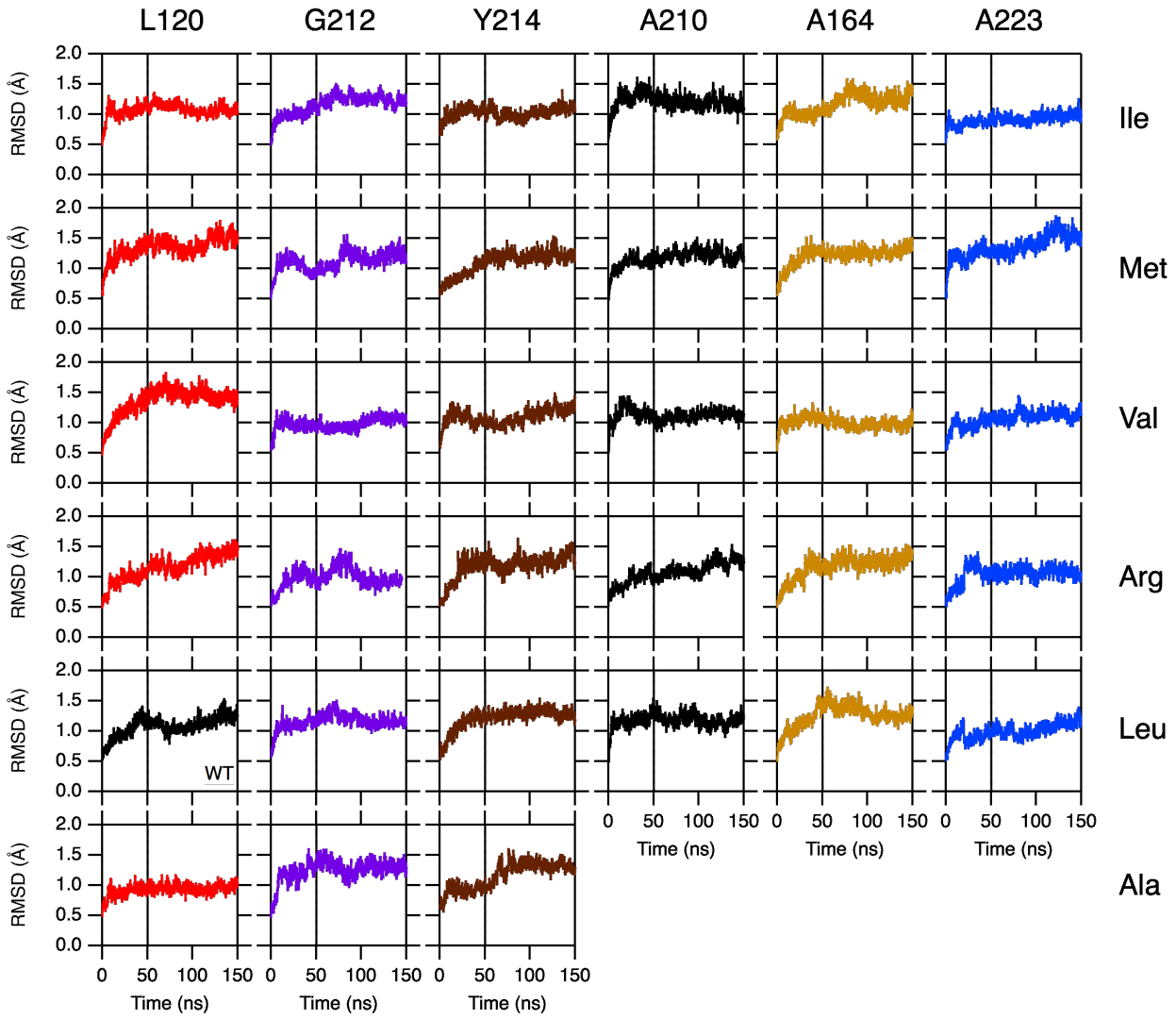
5.5 Acknowledgements

The authors would like to thank Dr. Sarah McDonald and Margo Goodall for assistance with cloning the variants used in this study, Mark Kreutzberger for growth and purification of some OmpLA variants used in this study, and Dr. Sarah McDonald and Dr. Patrick Fleming for experimental and computational guidance.

This work was funded by National Institutes of Health grants R01 GM079440 and T32 GM008403.

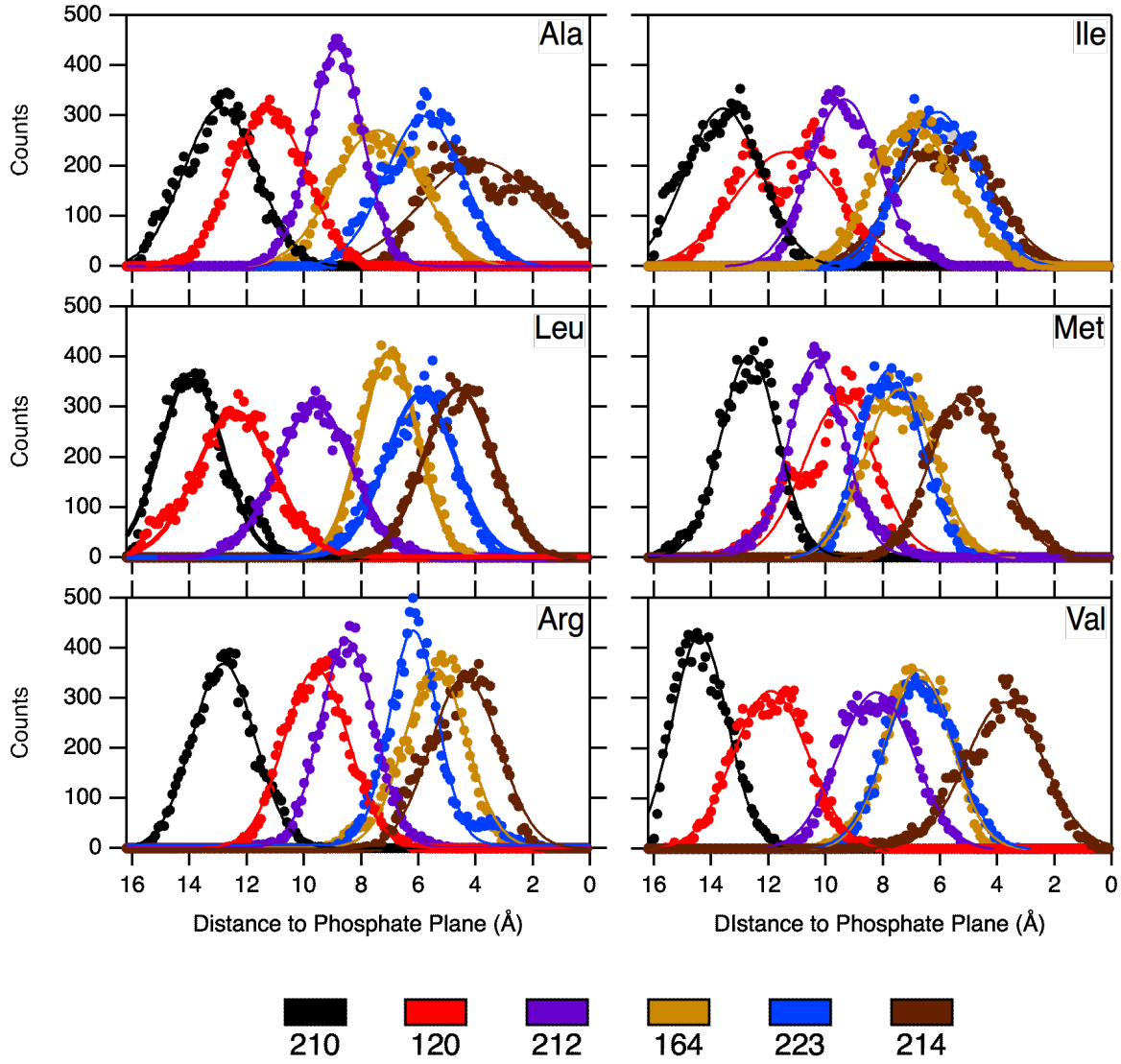
5.6 Figures

Figure 5.1 Molecular dynamics simulations of OmpLA variants are equilibrated after 50 ns.



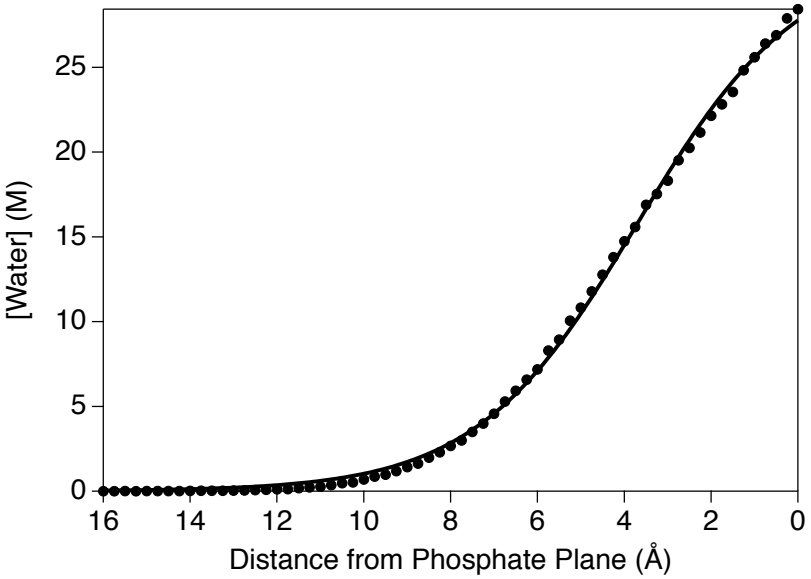
This figure shows the equilibration of each 150ns molecular dynamics trajectory by calculating the RMSD for the transmembrane beta-sheet of OmpLA as a function of time. RMSD plots are colored as in Figure 1. Further analysis of the trajectories was performed using the last 100ns of each trajectory.

Figure 5.2 Histograms of each side chain Cα position relative to the phosphate plane for each OmpLA variant.



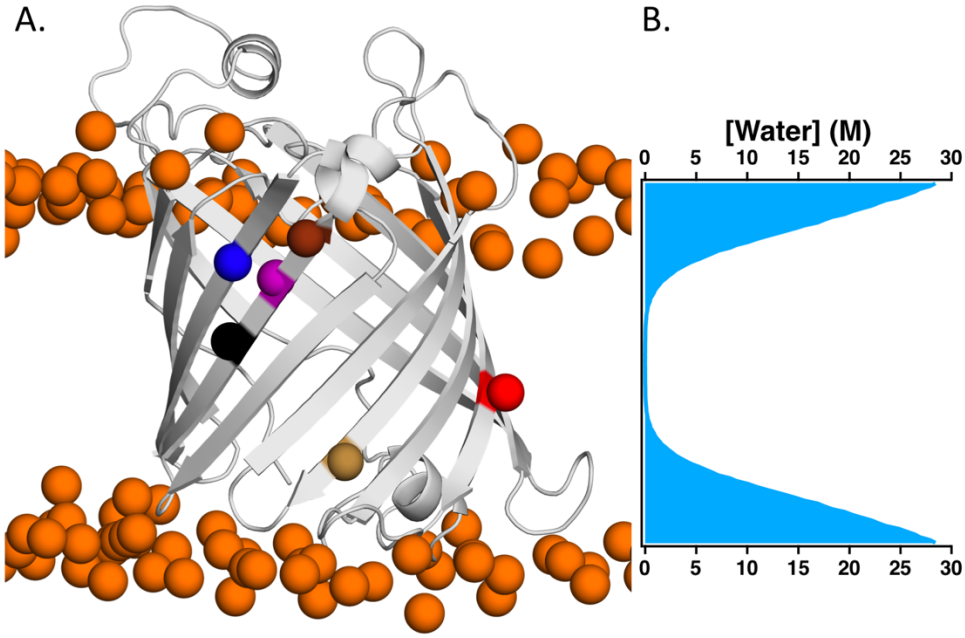
Cα positions for each variant were calculated for every time-step and binned by a 0.1Å step size over the last 100ns for each trajectory. Histograms relating the position of each variant in the bilayer relative to the phosphate plane are shown above. Solid lines are fits to a Gaussian distribution, from which the average position and error in the form of a standard deviation were extracted (Table 5.3). Site 223 overlaps in position with sites 164 and 214, which is why $\Delta\Delta G^{\circ}_{w,l}$ for Ile, Met, and Val were not measured at this site.

Figure 5.3 The Relationship Between [water] and Bilayer Z-Position is Well Described by a Sigmoid Function



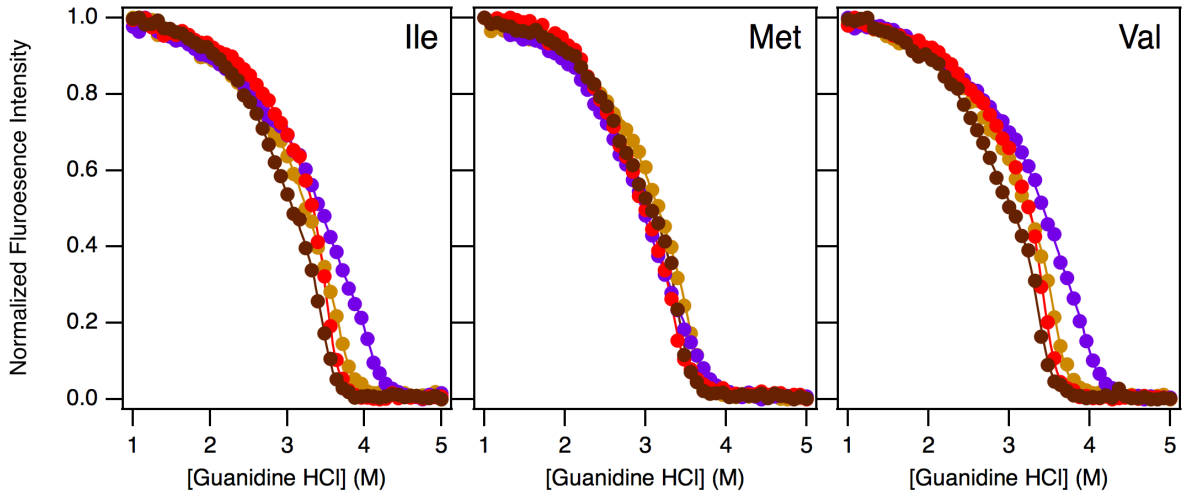
The data shown here (black circles) are derived from previously published molecular dynamics simulations of a neat DLPC bilayer using the density profile plugin (225). The data are well described by a single sigmoid indicated by the solid line. (Equation 5.4).

Figure 5.4 Host-sites on OmpLA exist in different water concentrations in the bilayer.



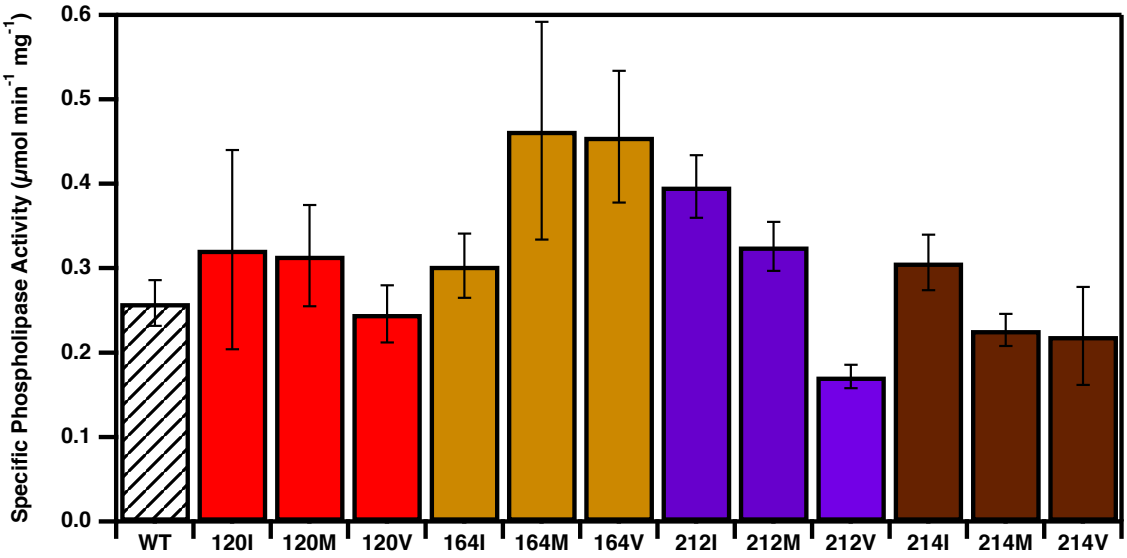
(A) A snapshot of a molecular dynamics simulation of WT OmpLA in a DLPC bilayer is shown. Phosphate atoms of the DLPC bilayer are colored orange and six host sites on OmpLA used in this study are shown as colored spheres (black:210, yellow:164, red:120, purple:212, blue:223, brown:214). (B) The gradient of water concentrations inside the phosphate plane of the bilayer is plotted as a function of position in the bilayer. These values were obtained from previously published simulations of neat DLPC bilayers (78). This water gradient is aligned with the structure of OmpLA at the two phosphate planes. The sites on OmpLA were chosen because they sample a wide range of z-positions positions and chemical compositions.

Figure 5.5 Representative folding titrations for Ile, Met, and Val OmpLA variants.



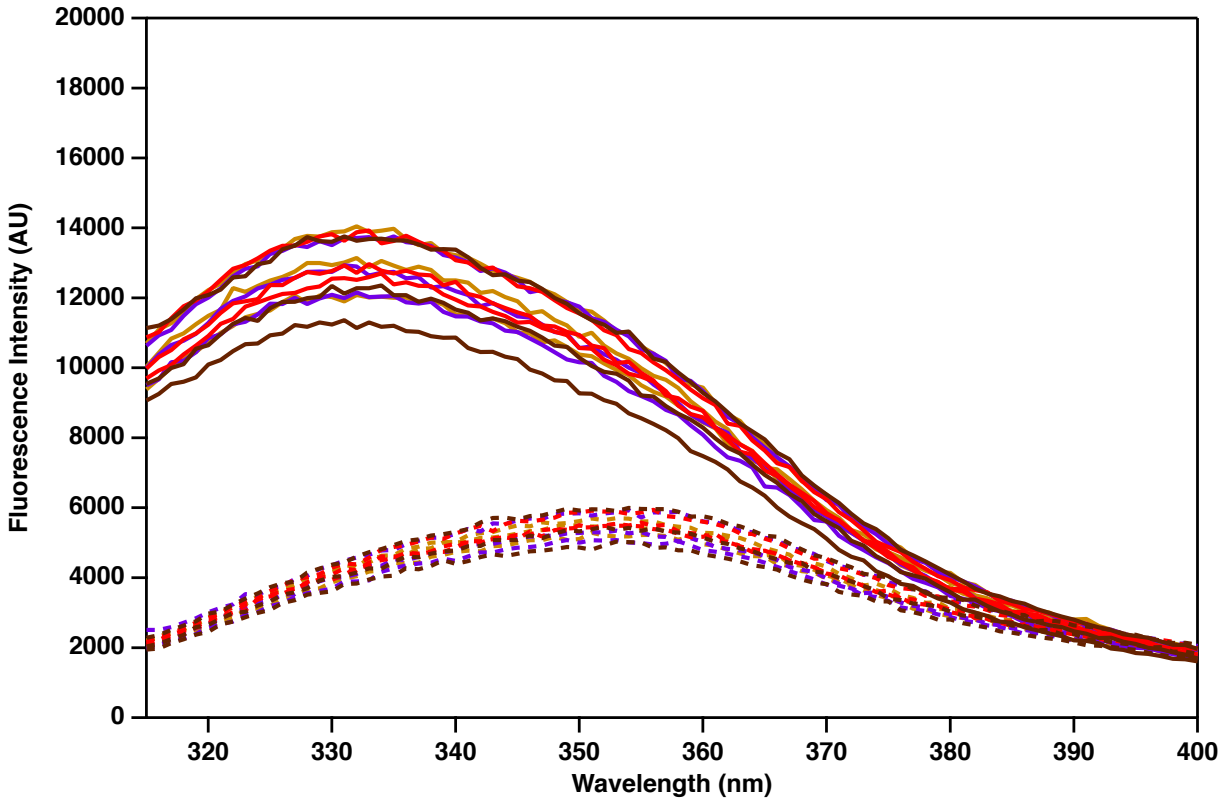
Representative chemical denaturation titrations are shown for each hydrophobic variant of OmpLA. Sites are colored as in Figure 5.4 (L120:red, A164:tan, G212:purple, Y214:brown). Fits to a three-state linear extrapolation model are shown as lines in the same color as the data set that they are fit to (75, 78, 97).

Figure 5.6 Nonpolar variants of OmpLA are enzymatically active.



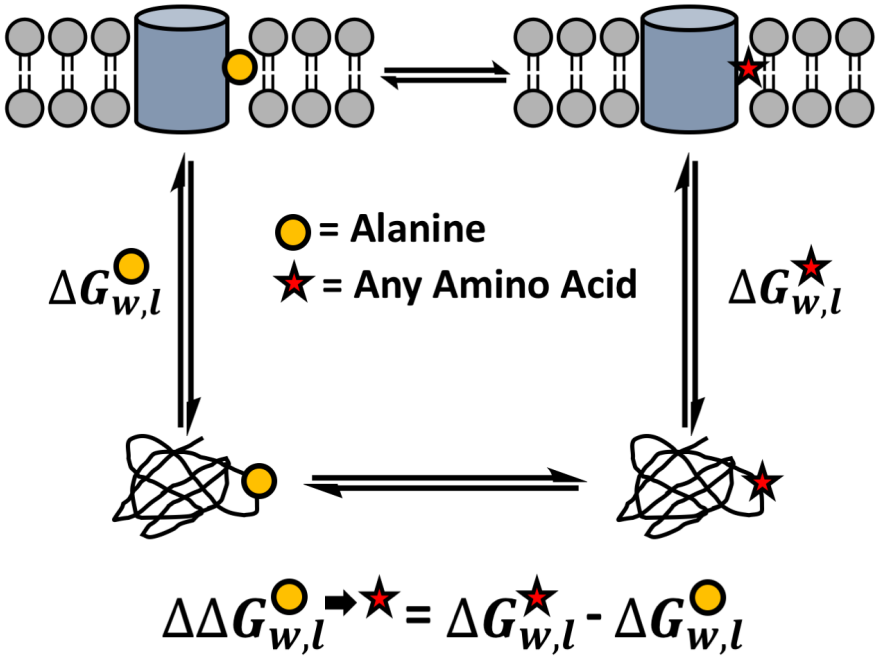
The activity of each new OmpLA variant was determined using a phospholipase activity assay on protein folded into DLPC LUVs in 1M Guanidine HCl (75, 78) The specific activities of each nonpolar guest-variant used in this study, and WT OmpLA, are shown colored as in Figure 5.4. Each variant has measurable enzymatic activity similar to WT, indicating that each variant is able to fold to a WT-like structure in DLPC bilayers. Error bars represent standard deviations from three independent trials.

Figure 5.7 Nonpolar Variants of OmpLA fold into DLPC LUVs.



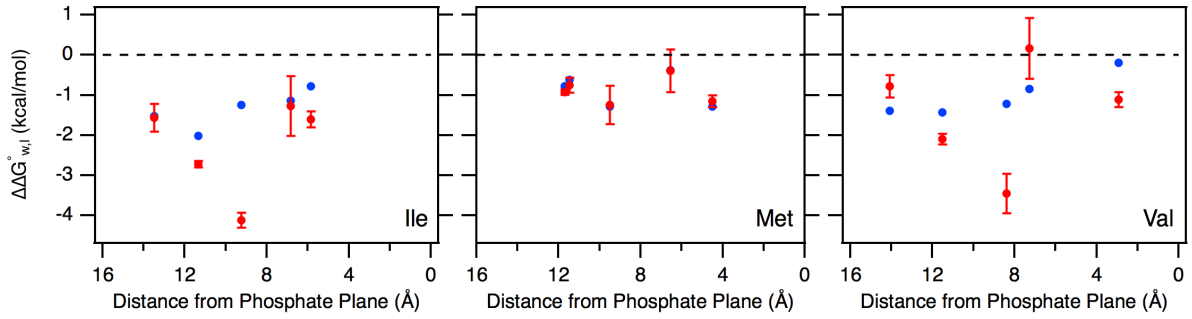
Tryptophan-fluorescence wavelength scans for OmpLA variants used in this study are shown. Solid lines correspond to folding conditions (1M Gdn) and dotted lines correspond to denaturing conditions (5M Gdn). The similarity of the shape and signal intensity for each OmpLA variant indicates that they all adopt similar structures in both folded and denatured conformations.

Figure 5.8 Schematic of the host-guest calculation of $\Delta\Delta G^{\circ}_{w,l}$.



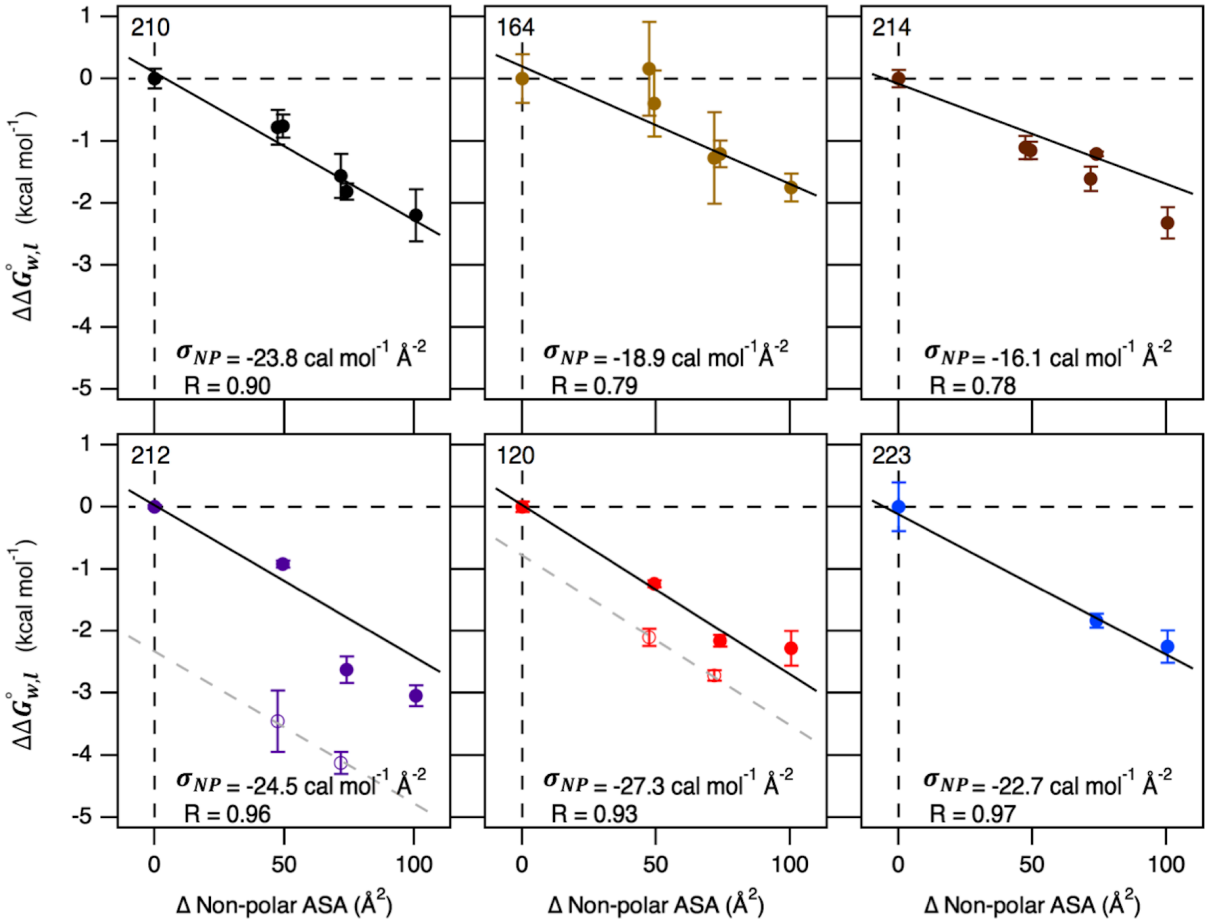
The stability ($\Delta G^{o}_{w,l}$) of both host (alanine) and guest (any other amino acid) variants at a site of OmpLA are measured using chemical denaturation titrations. The guest side chain transfer free energy with respect to alanine, $\Delta\Delta G^{o}_{w,l}$, is calculated by taking the difference in the host and guest $\Delta G^{o}_{w,l}$.

Figure 5.9 Comparison of Experimental and Computational $\Delta\Delta G^{\circ}_{w,l}$ for Ile, Met, and Val.



Comparison of experimentally measured $\Delta\Delta G^{o}_{w,l}$ for Ile, Met, and Val (red points) with computationally estimated $\Delta\Delta G^{o}_{w,l}$ (blue points) at these positions (226). Error bars for experimental data points correspond to standard deviations (Table 5.1).

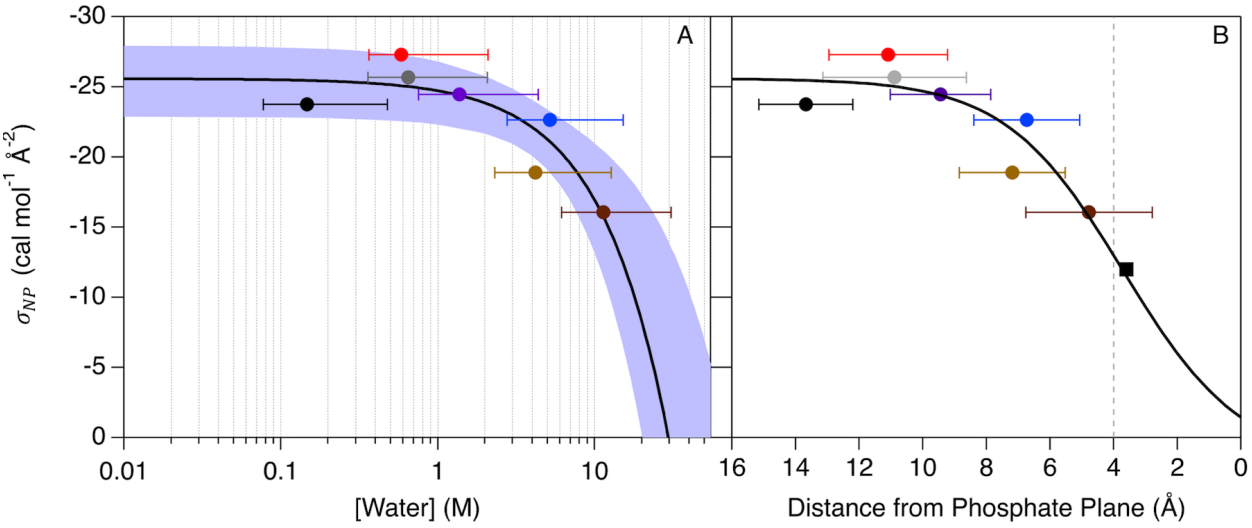
Figure 5.10 Nonpolar solvation parameter is not constant throughout the bilayer.



The nonpolar solvation parameter, σ_{NP}, was determined for each site on OmpLA by calculating the slope of the linear relationship between $\Delta\Delta G^{\circ}_{w,l}$ for each nonpolar side chain at that site with the change in buried surface area compared to alanine in a Gly-X-Gly peptide (Equation 5.1) (75). Linear fits were weighted by the error in $\Delta\Delta G^{\circ}_{w,l}$ (standard deviations found in Table 5.2) for each site, with the slope of the line (σ_{NP}) and the Pearson correlation coefficient reported for each fit shown at the bottom of each panel. For sites 212 and 120, beta-branched side chains isoleucine and valine were omitted from σ_{NP} determination as they had anomalously greater $\Delta\Delta G^{\circ}_{w,l}$ compared to the other

nonpolar side chains. The σ_{NP} determined from non-beta-branched side chains (solid line in 212 and 120 panels) fits the beta-branched side chains (dotted line), indicating that the increase in energy is derived from favorable local interactions that are restricted to beta-branched residues. By taking the difference in the y-intercept for the two lines, we estimate the energy gained due to local interactions for beta-branched residues at these sites to be 2.38 kcal mol^{-1} for site 212 and 0.80 kcal mol^{-1} for site 120. The σ_{NP} for site 223 was determined only using previously collected data (75).

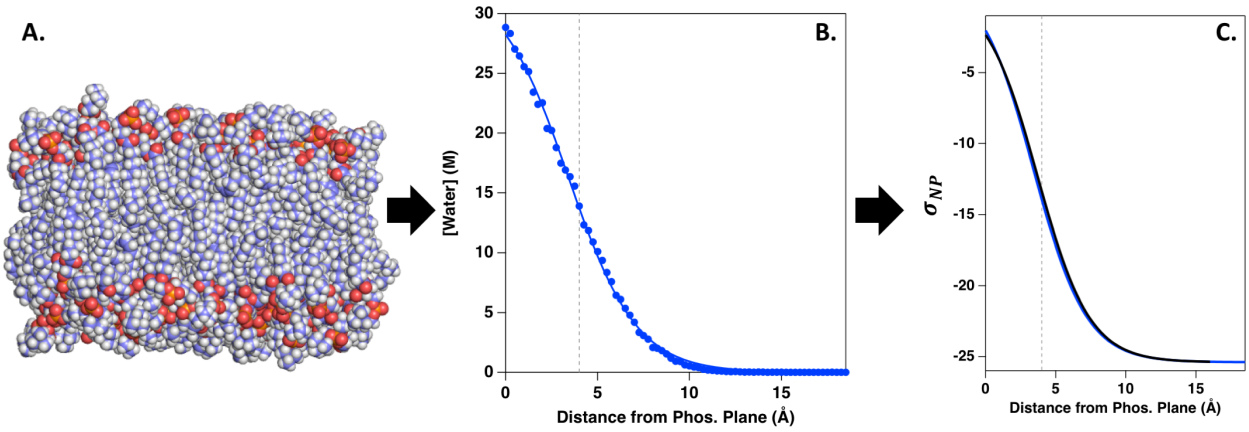
Figure 5.11 Nonpolar solvation parameter is linearly correlated with water concentration in the bilayer.



The water-to-bilayer σ_{NP} calculated at each host site in OmpLA and PagP plotted as a function of water concentration (Table 5.4) (**A**) and distance from the phosphate plane (**B**) (157). The σ_{NP} are colored according to the scheme in Figure 5.4, with the σ_{NP} derived from PagP position 111 colored gray. The position of σ_{NP} for a given site is the average of the positions of each side chain at that site and the error bars reflect the standard deviation (Table 5.1, NP row). σ_{NP} and water concentration are linearly correlated with the equation shown in the bottom right hand corner of Panel **A** (black line: $R^2 = 0.84$; 95% confidence interval shaded in light blue). Using the derived relationship between bilayer position and water concentration shown in Figure 5.3, the direct relationship between σ_{NP} and bilayer position can be determined (Panel **B**). We were also able to assign a previously measured "water-to-interface" solvation parameter (-12 cal mol^{-1} Å^{-2}) to an exact position in the bilayer (3.75 Å from the phosphate plane) using this function (shown as a black square in Panel **B**) and find that it is reporting on the position of the lipid

carbonyl plane in the bilayer (dashed vertical gray line) (99). The position-dependent function describing σ_{NP} allows for the ΔG°_{sc} of any nonpolar side chain to be determined for any position of the bilayer inside the phosphate plane (shown in Figure 5.13).

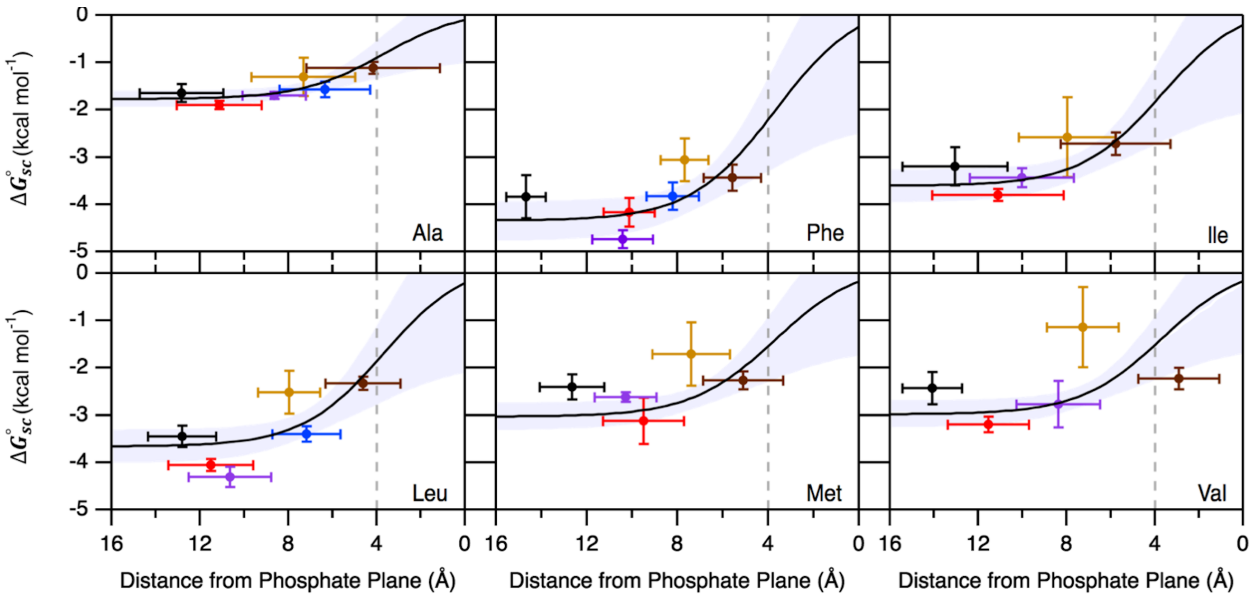
Figure 5.12 Flowchart for Applying $\sigma_{NP}(z)$ to Other Phospholipid Bilayers (Example of POPC)



This flowchart details how the relationship between the σ_{NP} and either water concentration or bilayer z-position can be applied to other bilayers using a POPC bilayer as an example. **(A)** First, the bilayer must be constructed properly and equilibrated using molecular dynamics simulations to ensure that the chemical composition of the bilayer interface is as accurate as possible. **(B)** Bilayer properties can be calculated using the density plug-in tool in VMD (225). To calculate the water density distribution, we calculate the average density of the oxygen atoms in the water molecules as a function of z-position using all frames from the 50ns simulation (blue circles). The relationship between [water] and *z*-position ($[water](z)$) can be fit to a sigmoid function (Equation 5.3; blue line). **(C)** Using the $[water](z)$ for the new bilayer (in this case POPC) and the $\sigma_{NP}([water])$ function derived in this paper (Figure 5.11, Equation 5.4), the $\sigma_{NP}(z)$ can be determined for a given bilayer. In Panel **C**, the black line is the $\sigma_{NP}(z)$ for the DLPC bilayer used in this study, and the blue line is the $\sigma_{NP}(z)$ for a POPC bilayer. We find that the $\sigma_{NP}(z)$ for both of

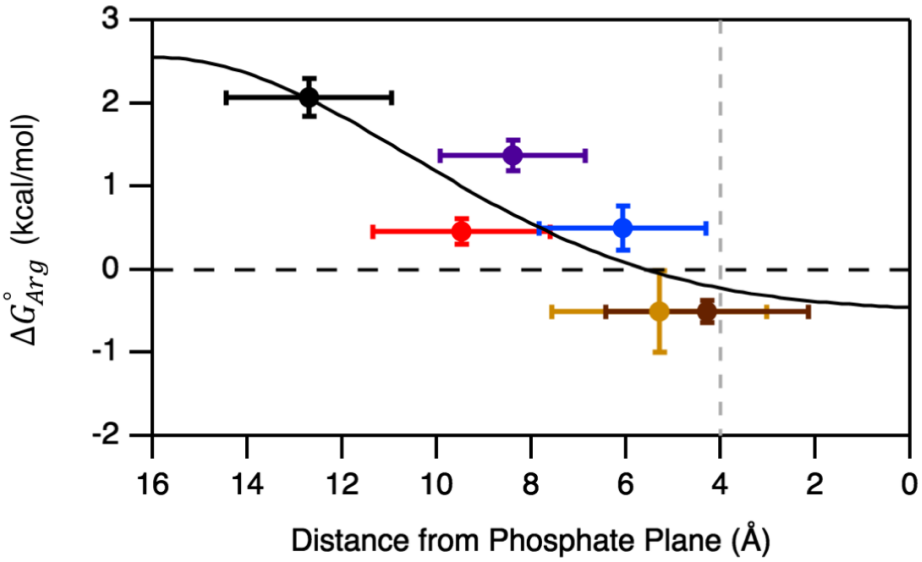
these bilayers are almost indistinguishable, indicating that $\sigma_{NP}(z)$ over the experimental range is not affected by changes in lipid acyl tail length or saturation.

Figure 5.13 Bilayer position dependence of nonpolar ΔG°_{sc}.



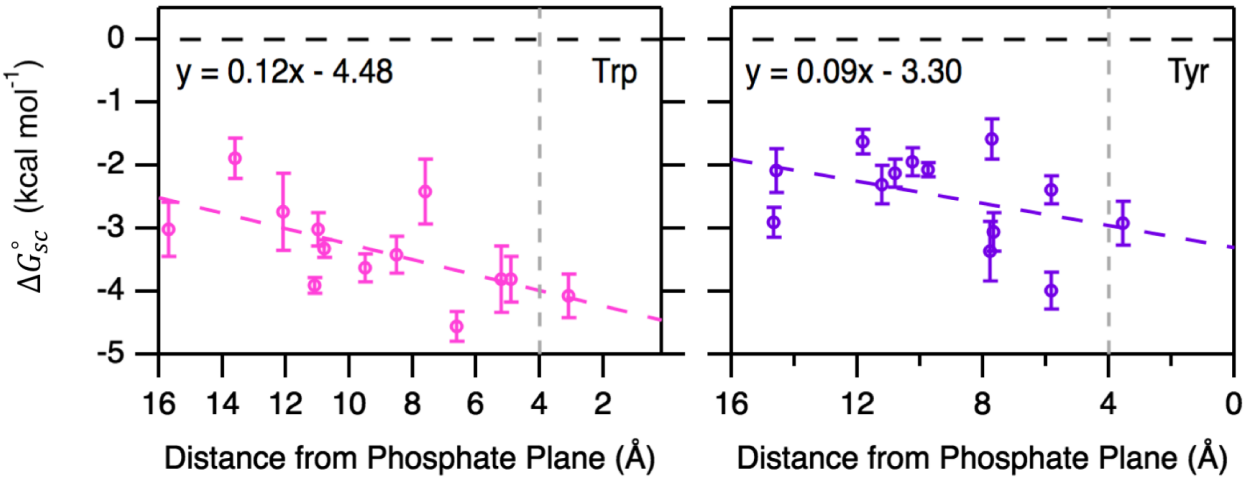
Reference-free side chain transfer free energies are plotted as a function of the average distance from the bilayer phosphate plane, as determined from molecular dynamics trajectories. ΔG°_{sc} for each host site in OmpLA are colored as in Figure 5.4. Error bars for both ΔG°_{sc} and bilayer position represent the standard deviations. The solid black line in each panel represents the simulated $\Delta G^{\circ}_{sc}(z)$ profile for each nonpolar side chain. This function is derived from $\sigma_{NP(z)}$ multiplied by the nonpolar surface area of each side chain (Equation 5.7). Vertical dotted lines represent the position of the lipid carbonyl groups.

Figure 5.14 Bilayer position dependence of arginine ΔG°_{sc}.



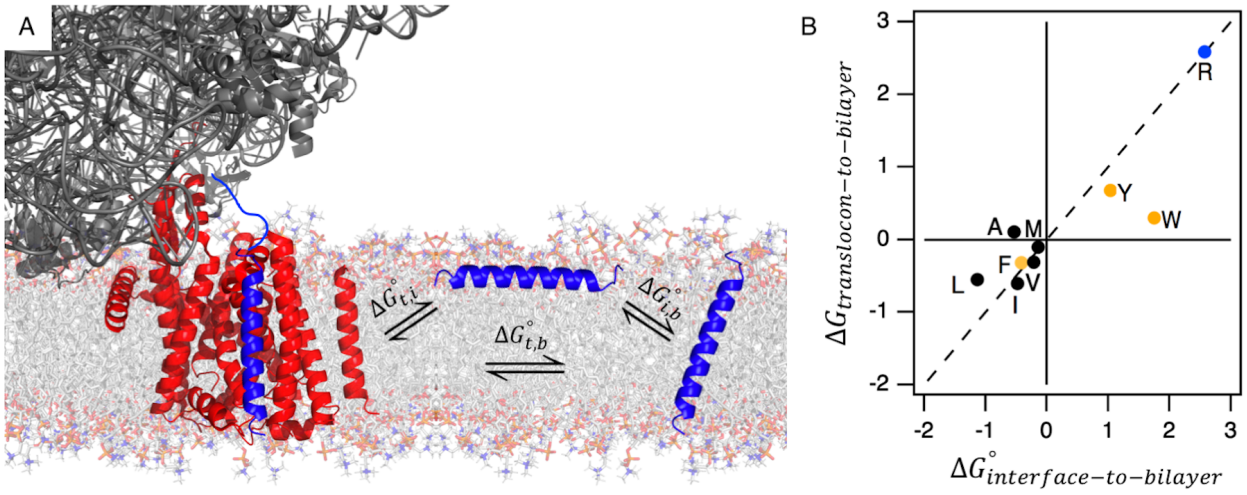
The ΔG°_{sc} for arginine at each of the six host sites on OmpLA are plotted as a function of their position relative to the phosphate plane in the DLPC bilayer (error bars represent standard deviations for both axes). We find the energetic penalty for inserting an arginine in the dehydrated core of the bilayer is approximately 2 kcal/mol. However, it is energetically favorable to have a lipid facing arginine within 6Å of the phosphate plane. We find that $\Delta G^{\circ}_{Arg}(z)$ can be defined by described by a Gaussian distribution (black line, Table 5.7). The description of the position dependence of ΔG°_{Arg} as a Gaussian function has been reported previously (217). The vertical, gray dotted dashed line references the position of the carbonyl plane in the bilayer relative to the phosphate plane.

Figure 5.15 Bilayer position dependence of aromatic $\Delta G°_{sc}$.



The reference-free ΔG^{o}_{sc} values for Trp and Tyr are shown as a function of bilayer position, with linear regression ΔG^{o}_{Trp} and ΔG^{o}_{Tyr} functions shown as dashed lines (R^2 = 0.38 and 0.20, respectively). Trp and Tyr are more favorable in the bilayer interface than the center of the bilayer, reflecting the preference of aromatic residues to be found in interfacial regions of transmembrane domains. The vertical, gray dotted dashed line references the position of the carbonyl plane in the bilayer relative to the phosphate plane.

Figure 5.14 The translocon energetically mimics the bilayer interface.



(A) A cartoon schematic of the co-translational insertion of a helix (blue) via the translocon (ribosome colored dark gray, translocon colored red; PDB code: 5GAE) is shown on the left. The helix can either partition to an interfacial conformation, which is energetically described by $\Delta G^\circ_{t,i}$, or to a transmembrane conformation, described by $\Delta G^\circ_{t,b}$ (*t* = translocon, *i* = interfacial, *b* = transmembrane). The interfacial-to-transmembrane conformation, which can occur in the absence of the translocon, is energetically described by $\Delta G^\circ_{i,b}$. **(B)** The values for $\Delta G^\circ_{t,b}$ and $\Delta G^\circ_{i,b}$ are plotted for nine side chains, with the dotted line representing energetic equivalence between the two equilibria (black = nonpolar, gold = aromatic, blue = ionizable) (15). For all side chains, except Trp, $\Delta G^\circ_{t,b} \approx \Delta G^\circ_{i,b}$, indicating that the translocon energetically mimics the interface ($\Delta G^\circ_{t,i} \approx 0$). The deviation for Trp is hypothesized to be due to the energetic preference of Trp to exist in the bilayer interface (78, 223).

5.7 Tables

Table 5.1 Average Cα positions for each OmpLA variant relative to the bilayer phosphate plane.

	Site on OmpLA					
	164	210	223	212	120	214
Ile	6.83 ± 2.02	13.50 ± 1.87	5.98 ± 1.90	9.25 ± 1.72	11.34 ± 2.65	5.84 ± 2.40
Leu	7.02 ± 1.41	13.94 ± 1.54	5.93 ± 1.73	9.63 ± 1.86	12.41 ± 1.91	4.55 ± 1.70
Met	7.40 ± 1.71	12.66 ± 1.41	7.73 ± 1.55	10.3 ± 1.37	9.50 ± 1.80	5.10 ± 1.75
Val	6.76 ± 1.62	14.40 ± 1.34	6.72 ± 1.73	8.23 ± 1.88	11.92 ± 1.84	3.74 ± 1.91
Ala	7.43 ± 2.18	12.92 ± 1.82	5.81 ± 1.92	8.86 ± 1.29	11.24 ± 1.84	3.85 ± 2.93
Phe	7.68 ± 1.06	14.68 ± 0.86	8.22 ± 1.17	10.41 ± 1.34	10.13 ± 1.13	5.57 ± 1.25
Arg	5.30 ± 1.54	12.71 ± 1.54	6.08 ± 1.19	8.39 ± 1.34	9.47 ± 1.54	4.29 ± 1.69
NP	7.19 ± 1.65	13.68 ± 1.48	6.73 ± 1.60	9.44 ± 1.54	11.09 ± 1.82	4.78 ± 1.95

Individual side chain average positions and standard deviations are determined from the Gaussian fits in Figure 5.2. The NP row is the average position and error for all nonpolar side chains (i.e. excluding arginine), which are used as the position and error for σ_{NP} plotted in Figure 5.13. All numbers are distances in angstroms from the average position of the phosphate plane in all-atom molecular dynamics simulations.

Table 5.2 Stabilities of alanine-dependent $\Delta\Delta G^{\circ}_{w,l}$ of Ile, Met, and Val OmpLA variants.

OmpLA Variant	$\Delta G^{o}_{w,l}$[a] (kcal mol^{-1})	$\Delta\Delta G^{o}_{w,l}$[b] (kcal mol^{-1})
120I	-30.91 ± 0.01	-2.72 ± 0.08
164I	-32.89 ± 0.63	-1.27 ± 0.74
212I	-36.55 ± 0.18	-4.12 ± 0.18
214I	-30.57 ± 0.14	-1.61 ± 0.20
120M	-29.43 ± 0.28	-1.24 ± 0.48
164M	-32.02 ± 0.36	-0.40 ± 0.53
212M	-33.35 ± 0.05	-0.92 ± 0.06
214M	-30.11 ± 0.03	-1.15 ± 0.14
120V	-30.29 ± 0.12	-2.10 ± 0.14
164V	-31.46 ± 0.64	0.16 ± 0.75
212V	-35.88 ± 0.49	-3.45 ± 0.49
214V	-30.07 ± 0.12	-1.11 ± 0.19

Errors are either standard deviations (n = 3) or standard error of the mean (n = 2)

[a]$\Delta G^{o}_{w,l}$ values are the Gibbs free energies for folding for each variant for the water (w)-to-lipid (l) transition. Chemical denaturation titrations were fit to a three state model with *m*-values held constant to previously determined values for each transition (2.03 and 7.18 kcal mol^{-1} M^{-1}, respectively) (75). $\Delta G^{o}_{w,l}$ above are the sum of the best-fit values for the two transitions in the three-state fit.

[b]$\Delta\Delta G^{o}_{w,l}$ values were calculated by subtracting the $\Delta G^{o}_{w,l}$ for the guest variant shown above from the $\Delta G^{o}_{w,l}$ for the previously reported alanine variant at each site (78). Error is propagated from the $\Delta G^{o}_{w,l}$ for all variants.

Table 5.3 Parameters for σ_{NP} Fits in Figure 5.10

Site	Slope (σ_{NP})	y-intercept
120	-27.3	0.03
164	-18.9	0.20
210	-23.8	0.10
212	-24.5	0.03
214	-16.1	-0.08
223	-22.7	-0.12

The slopes and intercepts for the weighted linear regressions shown in Figure 5.10 are reported here.

Table 5.4 Reference free ΔG°_{sc} for side chains at the six sites on OmpLA.

Site	Side Chain	ΔG^{o}_{sc} (kcal mol^{-1}) [a]	Site	Side Chain	ΔG^{o}_{sc} (kcal mol^{-1})[a]
120	A	-1.89 ± 0.09	164	A	-1.30 ± 0.41
	F	-4.17 ± 0.30		F	-3.05 ± 0.46
	I*	-3.80 ± 0.12		I	-2.58 ± 0.85
	L	-4.05 ± 0.13		L	-2.51 ± 0.46
	M	-3.12 ± 0.49		M	-1.71 ± 0.67
	R	0.46 ± 0.15		R	-0.50 ± 0.49
	V*	-3.19 ± 0.17		V	-1.14 ± 0.85
	W	-4.05 ± 0.13		W	-2.13 ± 0.51
	Y	-2.46 ± 0.30		Y	-3.06 ± 0.47
212	A	-1.69 ± 0.08	214	A	-1.11 ± 0.13
	F	-4.73 ± 0.19		F	-3.34 ± 0.28
	I*	-3.43 ± 0.20		I	-2.72 ± 0.24
	L	-4.31 ± 0.22		L	-2.32 ± 0.14
	M	-2.61 ± 0.10		M	-2.26 ± 0.19
	R	1.37 ± 0.18		R	-0.50 ± 0.13
	V*	-2.76 ± 0.49		V	-2.22 ± 0.23
	W	-3.28 ± 0.14		W	-3.77 ± 0.36
	Y	-2.05 ± 0.11		Y	-3.77 ± 0.29
223	A	-1.57 ± 0.16			
	F	-3.82 ± 0.29			
	I	*n/a*			
	L	-3.40 ± 0.17			
	M	*n/a*			
	R	0.50 ± 0.26			
	V	*n/a*			
	W	-4.67 ± 0.24			
	Y	-3.04 ± 0.31			

[a]ΔG^{o}_{sc} are the reference-free side chain transfer free energy for a given side chain at that position on OmpLA. The ΔG^{o}_{sc} for alanine are calculated by multiplying the nonpolar solvation parameter determined for that site (Figure S3) by the nonpolar surface area of an alanine side chain (69.1 Å^2). For all other side chains, ΔG^{o}_{sc} is calculated by adding the ΔG^{o}_{sc} for alanine to the $\Delta\Delta G^{\circ}_{w,l}$ measured for each side chain at that site on OmpLA (Table S1).

*ΔG_{sc}^{o} for beta-branched residues at sites 120 and 212 have been decreased by 0.80 and 2.38 kcal/mol respectively to reflect only the transfer free energy of these side chains and not local interactions.

Table 5.5 Local water concentration and σ_{NP} for each site on OmpLA and PagP.

Site	σ_{NP} (cal mol^{-1} Å^{-2})	Local [water] (M)
210	-23.8	0.15
120	-27.3	0.58
111 (PagP)	-25.7	0.65
212	-24.5	1.37
164	-18.9	4.20
223	-22.7	5.18
214	-16.1	11.33

The local [water] for each site were determined using the average nonpolar Cα position listed in Table 5.1 and using Equation 5.3. For site 111 on PagP, analyses were carried out using previously published simulations of WT PagP (157).

Table 5.6 Aromatic ΔG°_{sc} at seven additional sites on OmpLA.

	ΔG^{o}_{sc} (kcal mol^{-1}) [a]		
Site	Phe	Trp	Tyr
243	-3.92 ± 0.34	-4.07 ± 0.35	-2.91 ± 0.34
195	n/a	-3.62 ± 0.22	-1.94 ± 0.22
239	-4.74 ± 0.21	-2.73 ± 0.61	-2.13 ± 0.22
173	-4.29 ± 0.20	-3.02 ± 0.43	-2.90 ± 0.23
162	-3.44 ± 0.23	-3.02 ± 0.26	-1.63 ± 0.19
136	-3.43 ± 0.25	-3.41 ± 0.29	-1.58 ± 0.32
122	-4.27 ± 0.31	-3.80 ± 0.52	-2.39 ± 0.22

[a] ΔG^{o}_{sc} for these sites on OmpLA were determined using the predicted bilayer position dependent profile for ΔG^{o}_{Ala}. The bilayer positions of each variant were previously determined using molecular dynamics simulations, and the corresponding ΔG^{o}_{Ala} for alanine at that position in the bilayer was used to adjust the experimentally determined $\Delta\Delta G^{o}_{w,l}$ at these sites by McDonald and Fleming (78).

Table 5.7 $\Delta G_{sc}^{\circ}(z)$ Equations

Side Chain	Function
A	$\Delta G_{sc}^{\circ}(z) = 0.069 * \left(0.87 * \left(1.0 * 10^{-6} + \left\{ \frac{31.563}{1 + exp^{\frac{3.7015 - z}{-1.8604}}} \right\} \right) - 25.6 \right)$
F	$\Delta G_{sc}^{\circ}(z) = 0.170 * \left(0.87 * \left(1.0 * 10^{-6} + \left\{ \frac{31.563}{1 + exp^{\frac{3.7015 - z}{-1.8604}}} \right\} \right) - 25.6 \right)$
I	$\Delta G_{sc}^{\circ}(z) = 0.141 * \left(0.87 * \left(1.0 * 10^{-6} + \left\{ \frac{31.563}{1 + exp^{\frac{3.7015 - z}{-1.8604}}} \right\} \right) - 25.6 \right)$
L	$\Delta G_{sc}^{\circ}(z) = 0.143 * \left(0.87 * \left(1.0 * 10^{-6} + \left\{ \frac{31.563}{1 + exp^{\frac{3.7015 - z}{-1.8604}}} \right\} \right) - 25.6 \right)$
M	$\Delta G_{sc}^{\circ}(z) = 0.118 * \left(0.87 * \left(1.0 * 10^{-6} + \left\{ \frac{31.563}{1 + exp^{\frac{3.7015 - z}{-1.8604}}} \right\} \right) - 25.6 \right)$
R	$\Delta G_{sc}^{\circ}(z) = -0.5 + 3.05 exp^{\left\{ -\left(\frac{z-16}{7.75} \right)^2 \right\}}$
V	$\Delta G_{sc}^{\circ}(z) = 0.116 * \left(0.87 * \left(1.0 * 10^{-6} + \left\{ \frac{31.563}{1 + exp^{\frac{3.7015 - z}{-1.8604}}} \right\} \right) - 25.6 \right)$
W	$\Delta G_{sc}^{\circ}(z) = 0.12 * z - 4.48$
Y	$\Delta G_{sc}^{\circ}(z) = 0.09 * z - 3.30$

*z is z-position of Cα atom in the bilayer relative to the nearest phosphate plane

Chapter 6: Concluding Remarks

Efficient biogenesis of membrane proteins is essential for all forms of life. Experimental studies of membrane proteins are challenging due to the hydrophobic character of membrane proteins which makes them prone to irreversible aggregation. In my thesis work I have investigated important aspects of membrane protein biogenesis. In particular, I have advanced our understanding about both how the periplasmic chaperone SurA interacts with unfolded OMPs and how the effects of the local chemical environment affect side chain transfer free energies. Prior to the work presented here, little was known about either the apo-structure of the periplasmic chaperone SurA in solution or the structural details of how SurA binds and solubilizes unfolded client OMPs. Additionally, while side chain free energies are thought to dominate the thermodynamic stability of membrane proteins, the effect of the local environment on these thermodynamic parameters had not been explicitly established. My thesis work provides important steps forward and raises exciting new questions about membrane protein biogenesis and folding.

SurA populates a diverse conformational ensemble

We found that SurA adopts a variety of conformations in solution through a combination of equilibrium thermodynamics, scattering, and computational modeling (150). The conformational ensemble reveals that the two PPIase domains of SurA compete for interaction with the core domain, and that an “open” conformation of SurA in which all domains are structurally isolated is populated in solution. Serendipitously, our findings were corroborated by two additional, orthogonal studies from other groups that were published within months of my work (69, 70). The intrinsic formation of the “open”

conformation SurA was an important finding given the recent evidence that this conformation is the chaperone active conformation of SurA (69, 70, 108). Questions remain about the mechanism of uOMP recognition by SurA and how this initial capture complex is coupled to domain dynamics. The findings presented here provide the framework to ascertain whether SurA functions via an induced-fit model, where initial uOMP interaction favors a switch to the open conformation of SurA, or conformational selection, where SurA can only bind uOMPs when already in the open conformation.

Additionally, the connection between the apparent autoinhibition of SurA chaperone function is not well understood at this time. The proposed conformational ensemble minimizes the exposure of the uOMP binding groove, indicating that exposing the binding groove could result in unfavorable interactions between SurA and non-native binding partners in the periplasm. This is opposed by studies that show that deletion of one or both PPIase domains does not induce an observable phenotype in E. *coli* cells (40, 57, 65). To understand the apparent auto-inhibition of SurA chaperone activity, a better understanding of exactly how SurA binds and solubilizes client uOMPs is necessary.

Expansion of uOMPs upon binding to SurA

Another key finding in our studies of SurA is that it greatly expands client uOMPs upon binding (108). Surprisingly, the size of the SurA-uOMP complex is approximately equal to the estimated width of periplasm. This expansion was measured *in vitro* in dilute conditions which raises questions as to whether uOMPs are expanded by SurA to the same degree *in vivo*. The periplasm is a crowded environment and contains the

peptidoglycan layer, both of which may limit the expansion of uOMPs by SurA. Future studies are required to assess the extent of uOMP expansion by SurA in various crowded environments.

In addition to the extent of expansion in the periplasm, the presence of other uOMP chaperone proteins presents the possibility of the formation of heterocomplexes with both SurA and another chaperone bound to a client uOMP at the same time. Of particular interest is BamA, which is the central component of the BAM complex. SurA has been shown to interact with BamA, but the structure of this interaction is not yet resolved. It is unclear whether SurA intrinsically interacts with BamA or if uOMP is required to promote an interaction, as SurA is proposed to hand off uOMP clients to the BAM complex. The possibility of a SurA-uOMP-BAM interaction has been recently complicated by the determination of a structure connecting the translocon and the BAM complex by cryo-EM (45). It is unclear at this point how SurA is involved in this periplasm bridge for uOMPs, though the authors include SurA in their mechanistic model. The methodologies presented here, particularly the combination of binding thermodynamics and XL-MS, should provide the basis to assess the possibility of uOMP handoff between chaperones in the periplasm.

Local Protein Structure Does Not Affect ΔG°_{sc}

In addition to studying that solubilization of unfolded membrane proteins in aqueous solutions, I also studied membrane protein folding. Side chain transfer free energies have been measured in a variety of contexts: side chain analogs, small peptides, and whole proteins partitioning into either organic solvents or phospholipid bilayers.

Despite the wealth of studies aiming to quantify side chain transfer free energies, neither the role of nearest-neighbor side chains nor the effect of varying protein architectures in modulating these values have been investigated. I addressed these questions by measuring the side chain transfer free energies at a single site on the protein scaffold PagP (157). Previous work in the Fleming lab had focused solely on measuring side chain transfer free energies using OmpLA as the protein scaffold (75, 78). By changing the protein scaffold and picking host sites that reside in chemically identical locations within the bilayer, I was able to begin to understand how protein structure and sequence affect side chain transfer free energies.

On the whole, I found that side-chain transfer free energies were relatively unaffected by the protein scaffold. This finding was encouraging for future measurements of side chain transfer free energies as it provides the basis for applying side chain transfer free energies ubiquitously. The similarity of the measured transfer energetics given the structural differences between OmpLA and PagP (strand number and tilt) indicate that side chain transfer free energies are relatively independent of structure of a particular protein. This is crucial for the application of these values to alpha-helical proteins, which can vary dramatically in tilt angle with respect to the membrane.

Additionally, this study has implications for predicting the stability of membrane protein TMDs, which is one of the predominant uses of these energetic terms currently. Because most side chains had similar transfer free energies in both PagP and OmpLA, the impact of nearest-neighbor side chains on modulating the magnitude of the transfer free energies of a given side chain seems to be negligible. This implies that there may be

little influence of neighboring residues on a TMD on the transfer free energy of a given side chain. The biggest caveat to this finding is that substantial energetic cooperativity was shown for Arg, as inserting two Arg into a TMD had less of an energetic penalty than the inclusion of each side chain independently (75). Interactions between neighboring aromatic residues have also been observed, leading to non-additive transfer free energies (78). Overall, the cumulative findings suggest that cooperativity may only be energetically substantial when aromatic or polar side chains are found in a TMD. Applying the experimental approach used by Moon and Fleming to assess the cooperativity of other side chain pairs is an important next step in understanding membrane protein stability.

Local Bilayer Hydrophobicity Modulates Nonpolar ΔG_{sc}°

While the importance of cooperativity for nonpolar side chains remains unclear, their transfer free energy is modulated by their position in the bilayer. I measured the bilayer-position dependence of nonpolar side chains and correlated the measured transfer free energies with the polarity gradient found across the bilayer interface (239). My work on measuring the bilayer position dependence of side chain transfer free energies provides a strong basis for future studies of this nature. The most straightforward next step for this project would be to continue measuring the bilayer *z*-position dependence of side chain transfer free energies and creating a matrix relating side chain position and transfer free energy. Currently, the Fleming lab has measured the depth dependence of nine of the twenty side chains, leaving ample room for future studies. The remaining side chains are generally polar in nature, with serine and threonine being of particular interest. These two side chains are small enough that they are not expected to

be able to snorkel to the bilayer interface like arginine (77). Thus, understanding the position dependence of the transfer free energy for these two side chains in the bilayer will potentially allow for us to estimate the cost of satisfying a hydrogen bond across the bilayer.

Finishing the “matrix” will allow for a more full comparison with published estimates of side chain transfer free energies across the bilayer using other systems and allow us to begin to parse out the differences between the different experiment types (205, 217). This type of meta-analysis will allow for a more holistic understanding of the role of side chain partitioning in membrane protein folding to be better understood. Combined with a more complete understanding of the influence of neighboring side chains on the transfer free energy of a given side chain, a complete understanding of the role that individual side chains play in membrane protein folding and stability seems attainable. Further work elucidating the thermodynamic cooperativity of the insertion of multiple side chains would be necessary to apply this construct accurately to estimate whole-protein stabilities. I think a lofty, long-term goal for this project would be to develop a software package that incorporates bilayer *z*-position dependent information and the identity and position of neighboring residues to provide accurate estimates of membrane protein stability. This could be applied to estimating the consequences of disease-causing mutations on the stability of membrane proteins or elucidate the energetic differences between different conformations of membrane proteins.

References

1. C. M. Dobson, Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
2. K. Vinothkumar, R. Henderson, Structures of membrane proteins. *Q. Rev. Biophys.* **43** (2010).
3. W. Kühlbrandt, Structure and Mechanisms of F-Type ATP Synthases. *Annu. Rev. Biochem.* **88**, 515–549 (2019).
4. W. Palm, C. B. Thompson, Nutrient acquisition strategies of mammalian cells. *Nature* **546**, 234–242 (2017).
5. C. R. H. Raetz, C. M. Reynolds, M. S. Trent, R. E. Bishop, Lipid a modification systems in gram-negative bacteria. *Annu. Rev. Biochem.* **76**, 295–329 (2007).
6. S. W. Englander, L. Mayne, The nature of protein folding pathways. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15873–15880 (2014).
7. N. C. Bauer, P. W. Doetsch, A. H. Corbett, Mechanisms Regulating Protein Localization. *Traffic* **16**, 1039–1061 (2015).
8. Z. Sun, J. L. Brodsky, Protein quality control in the secretory pathway. *J. Cell Biol.* **218**, 3171–3187 (2019).
9. A. Ruggiano, O. Foresti, P. Carvalho, ER-associated degradation: Protein quality control and beyond. *J. Cell Biol.* **204**, 869–879 (2014).
10. A. Guna, R. S. Hegde, Transmembrane Domain Recognition during Membrane Protein Biogenesis and Quality Control. *Curr. Biol.* **28**, R498–R511 (2018).
11. P. Walter, A. E. Johnson, Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* **10**, 87–119 (1994).

12. K. Denks, *et al.*, The Sec translocon mediated protein transport in prokaryotes and eukaryotes. *Mol. Membr. Biol.* **31**, 58–84 (2014).

13. T. A. Rapoport, L. Li, E. Park, Structural and mechanistic insights into protein translocation. *Annu. Rev. Cell Dev. Biol.* **33**, 369–390 (2017).

14. K. Ito, N. Shimokawa-Chiba, S. Chiba, Sec translocon has an insertase-like function in addition to polypeptide conduction through the channel. *F1000Research* **8** (2019).

15. T. Hessa, *et al.*, Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377–81 (2005).

16. F. Cymer, G. Von Heijne, S. H. White, Mechanisms of integral membrane protein insertion and folding. *J. Mol. Biol.* **427**, 999–1022 (2015).

17. S. Shao, R. S. Hegde, Membrane protein insertion at the endoplasmic reticulum. *Annu. Rev. Cell Dev. Biol.* **27**, 25–56 (2011).

18. D. Chaturvedi, R. Mahalakshmi, Transmembrane β-barrels: Evolution, folding and energetics. *Biochim. Biophys. acta. Biomembr.* **1859**, 2467–2482 (2017).

19. K. Weinhäupl, *et al.*, Structural Basis of Membrane Protein Chaperoning through the Mitochondrial Intermembrane Space. *Cell* **175**, 1365-1379.e25 (2018).

20. A. I. C. Höhr, *et al.*, Membrane protein insertion through a mitochondrial β-barrel gate. *Science (80-.).* **359** (2018).

21. A. M. Glauert, M. J. Thornley, The topography of the bacterial cell wall. *Annu. Rev. Microbiol.* **23**, 159–198 (1969).

22. C. Sohlenkamp, O. Geiger, Bacterial membrane lipids: Diversity in structures and pathways. *FEMS Microbiol. Rev.* **40**, 133–159 (2015).

23. S. J. Facey, A. Kuhn, Membrane integration of E. coli model membrane proteins. *Biochim.*

Biophys. Acta - Mol. Cell Res. **1694**, 55–66 (2004).

24. A. M. Plummer, K. G. Fleming, From Chaperones to the Membrane with a BAM! *Trends Biochem Sci* **41**, 872–882 (2016).

25. R. Koebnik, K. P. Locher, P. Van Gelder, Structure and function of bacterial outer membrane proteins: Barrels in a nutshell. *Mol. Microbiol.* **37**, 239–253 (2000).

26. H. J. Snijder, *et al.*, Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature* **401**, 717–721 (1999).

27. A. Baslé, G. Rummel, P. Storici, J. P. Rosenbusch, T. Schirmer, Crystal Structure of Osmoporin OmpC from E. coli at 2.0 Å. *J. Mol. Biol.* **362**, 933–942 (2006).

28. E. Yamashita, M. V. Zhalnina, S. D. Zakharov, O. Sharma, W. A. Cramer, Crystal structures of the OmpF porin: Function in a colicin translocon. *EMBO J.* **27**, 2171–2180 (2008).

29. M. Pazos, K. Peters, Peptidoglycan. *Subcell. Biochem.* **92**, 127–168 (2019).

30. M. Babu, *et al.*, Global landscape of cell envelope protein complexes in Escherichia coli. *Nat. Biotechnol.* **36**, 103–112 (2018).

31. W. Vollmer, D. Blanot, M. A. De Pedro, Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* **32**, 149–167 (2008).

32. H. J. Lessen, P. J. Fleming, K. G. Fleming, A. J. Sodt, Building Blocks of the Outer Membrane: Calculating a General Elastic Energy Model for β-Barrel Membrane Proteins. *J. Chem. Theory Comput.* **14**, 4487–4497 (2018).

33. E. R. Rojas, *et al.*, The outer membrane is an essential load-bearing element in Gram-negative bacteria. *Nature* **559**, 617–621 (2018).

34. F. Stull, J.-M. Betton, J. C. A. Bardwell, Periplasmic Chaperones and Prolyl Isomerases. *EcoSal Plus* **8** (2018).

35. A. Sala, P. Bordes, P. Genevaux, Multitasking SecB chaperones in bacteria. *Front. Microbiol.* **5**, 666 (2014).

36. D. Schibich, *et al.*, Global profiling of SRP interaction with nascent polypeptides. *Nature* **536**, 219–223 (2016).

37. T. Cranford-Smith, D. Huber, The way is the goal: How SecA transports proteins across the cytoplasmic membrane in bacteria. *FEMS Microbiol. Lett.* **365** (2018).

38. M. Paetzel, Structure and mechanism of Escherichia coli type I signal peptidase. *Biochim. Biophys. Acta - Mol. Cell Res.* **1843**, 1497–1508 (2014).

39. S. E. Rollauer, M. A. Sooreshjani, N. Noinaj, S. K. Buchanan, Outer membrane protein biogenesis in Gram-negative bacteria. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370** (2015).

40. G. R. Soltes, J. Schwalm, D. P. Ricci, T. J. Silhavy, The activity of Escherichia coli chaperone SurA is regulated by conformational changes involving a parvulin domain. *J. Bacteriol.* **198**, 921–929 (2016).

41. Z. Chang, The function of the DegP (HtrA) protein: Protease versus chaperone. *IUBMB Life* **68**, 904–907 (2016).

42. R. Wu, R. Stephenson, A. Gichaba, N. Noinaj, The big BAM theory: An open and closed case? *Biochim. Biophys. Acta - Biomembr.* **1862** (2020).

43. N. Noinaj, S. E. Rollauer, S. K. Buchanan, The β-barrel membrane protein insertase machinery from Gram-negative bacteria. *Curr. Opin. Struct. Biol.* **31**, 35–42 (2015).

44. D. J. Sherman, *et al.*, Lipopolysaccharide is transported to the cell surface by a membrane-Tomembrane protein bridge. *Science (80-.).* **359**, 798–801 (2018).

45. S. Alvira, *et al.*, Inter-membrane association of the Sec and BAM translocons for bacterial outer-membrane biogenesis. *Elife* **9** (2020).

46. A. R. Ureta, R. G. Endres, N. S. Wingreen, T. J. Silhavy, Kinetic analysis of the assembly of the outer membrane protein LamB in Escherichia coli mutants each lacking a secretion or targeting factor in a different cellular compartment. *J. Bacteriol.* **189**, 446–454 (2007).

47. S. M. Costello, A. M. Plummer, P. J. Fleming, K. G. Fleming, Dynamic periplasmic chaperone reservoir facilitates biogenesis of outer membrane proteins. *Proc Natl Acad Sci U S A* **113**, E4794-800 (2016).

48. G. Mas, S. Hiller, Conformational plasticity of molecular chaperones involved in periplasmic and outer membrane protein folding. *FEMS Microbiol. Lett.* **365** (2018).

49. G. Mas, J. Thoma, S. Hiller, The Periplasmic Chaperones Skp and SurA. *Subcell. Biochem.* **92**, 169–186 (2019).

50. A. P. Chum, S. R. Shoemaker, P. J. Fleming, K. G. Fleming, Plasticity and transient binding are key ingredients of the periplasmic chaperone network. *Protein Sci.* **28**, 1340–1349 (2019).

51. K. Denoncin, J. Schwalm, D. Vertommen, T. J. Silhavy, J.-F. Collet, Dissecting the Escherichia coli periplasmic chaperone network using differential proteomics. *Proteomics* **12**, 1391–401 (2012).

52. S. W. Lazar, M. Almirón, A. Tormo, R. Kolter, Role of the Escherichia coli SurA protein in stationary-phase survival. *J. Bacteriol.* **180**, 5704–11 (1998).

53. J. G. Sklar, T. Wu, D. Kahne, T. J. Silhavy, Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli. *Genes Dev.* **21**, 2473–2484 (2007).

54. A. E. Rizzitello, J. R. Harper, T. J. Silhavy, Genetic evidence for parallel pathways of chaperone activity in the periplasm of Escherichia coli. *J. Bacteriol.* **183**, 6794–6800 (2001).

55. G. Hennecke, J. Nolte, R. Volkmer-Engert, J. Schneider-Mergener, S. Behrens, The periplasmic chaperone SurA exploits two features characteristic of integral outer membrane proteins for selective substrate recognition. *J. Biol. Chem.* **280**, 23540–23548 (2005).

56. D. Vertommen, N. Ruiz, P. Leverrier, T. J. Silhavy, J.-F. Collet, Characterization of the role of the Escherichia coli periplasmic chaperone SurA using differential proteomics. *Proteomics* **9**, 2432–43 (2009).

57. J. R. Humes, *et al.*, The Role of SurA PPIase Domains in Preventing Aggregation of the Outer-Membrane Proteins tOmpA and OmpT. *J. Mol. Biol.* **431**, 1267–1283 (2019).

58. H. M. Webb, L. W. Ruddock, R. J. Marchant, K. Jonas, P. Klappa, Interaction of the periplasmic peptidylprolyl cis-trans isomerase SurA with model peptides: The N-terminal region of SurA is essential and sufficient for peptide binding. *J. Biol. Chem.* **276**, 45622–45627 (2001).

59. S. Behrens, R. Maier, H. de Cock, F. X. Schmid, C. A. Gross, The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. *EMBO J.* **20**, 285–94 (2001).

60. J. Thoma, B. M. Burmann, S. Hiller, D. J. Müller, Impact of holdase chaperones Skp and SurA on the folding of β-barrel outer-membrane proteins. *Nat. Struct. Mol. Biol.* **advance on**, 795–802 (2015).

61. G. Li, *et al.*, Single-Molecule Detection Reveals Different Roles of Skp and SurA as Chaperones. *ACS Chem. Biol.* **13**, 1082–1089 (2018).

62. E. Bitto, D. B. McKay, Binding of phage-display-selected peptides to the periplasmic chaperone protein SurA mimics binding of unfolded outer membrane proteins. *FEBS Lett.* **568**, 94–98 (2004).

63. E. Bitto, D. B. McKay, The periplasmic molecular chaperone protein SurA binds a peptide motif that is characteristic of integral outer membrane proteins. *J. Biol. Chem.* **278**, 49316–49322 (2003).

64. K. H. Stymest, P. Klappa, The periplasmic peptidyl prolyl cis-trans isomerases PpiD and SurA have partially overlapping substrate specificities. *FEBS J.* **275**, 3470–3479 (2008).

65. D. P. Ricci, J. Schwalm, M. Gonzales-Cope, T. J. Silhavy, The activity and specificity of the outer membrane protein chaperone SurA are modulated by a proline isomerase domain. *MBio* **4** (2013).

66. Y. Wang, *et al.*, A supercomplex spanning the inner and outer membranes mediates the biogenesis of β-barrel outer membrane proteins in bacteria. *J. Biol. Chem.* **291**, 16720–16729 (2016).

67. E. Bitto, D. B. McKay, Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure* **10**, 1489–1498 (2002).

68. X. Xu, S. Wang, Y. X. Hu, D. B. McKay, The Periplasmic Bacterial Molecular Chaperone SurA Adapts its Structure to Bind Peptides in Different Conformations to Assert a Sequence Preference for Aromatic Residues. *J. Mol. Biol.* **373**, 367–381 (2007).

69. A. N. Calabrese, *et al.*, Inter-domain dynamics in the chaperone SurA and multi-site binding to its outer membrane protein clients. *Nat. Commun.* **11**, 1–16 (2020).

70. M. Jia, *et al.*, Conformational Dynamics of the Periplasmic Chaperone SurA. *Biochemistry* **59**, 3235–3246 (2020).

71. G. Mas, *et al.*, Regulation of chaperone function by coupled folding and oligomerization. *Sci. Adv.* **6** (2020).

72. S. Pan, C. Yang, X. S. Zhao, Affinity of Skp to OmpC revealed by single-molecule

detection. *Sci. Rep.* **10** (2020).

73. J. L. Popot, D. M. Engelman, Membrane Protein Folding and Oligomerization: The Two-Stage Model. *Biochemistry* **29**, 4031–4037 (1990).

74. G. H. M. Huysmans, S. a Baldwin, D. J. Brockwell, S. E. Radford, The transition state for folding of an outer membrane protein. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4099–4104 (2010).

75. C. P. Moon, K. G. Fleming, Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10174–7 (2011).

76. C. P. Moon, N. R. Zaccai, P. J. Fleming, D. Gessmann, K. G. Fleming, Membrane protein thermodynamic stability may serve as the energy sink for sorting in the periplasm. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 4285–90 (2013).

77. P. J. Fleming, J. A. Freites, C. P. Moon, D. J. Tobias, K. G. Fleming, Outer membrane phospholipase A in phospholipid bilayers: A model system for concerted computational and experimental investigations of amino acid side chain partitioning into lipid bilayers. *Biochim. Biophys. Acta - Biomembr.* **1818**, 126–134 (2012).

78. S. K. McDonald, K. G. Fleming, Aromatic Side Chain Water-to-Lipid Transfer Free Energies Show a Depth Dependence across the Membrane Normal. *J. Am. Chem. Soc.* **138**, 7946–7950 (2016).

79. H. J. Lessen, A. Majumdar, K. G. Fleming, Backbone Hydrogen Bond Energies in Membrane Proteins Are Insensitive to Large Changes in Local Water Concentration. *J. Am. Chem. Soc.* **142**, 6227–6235 (2020).

80. Z. Cao, J. M. Hutchison, C. R. Sanders, J. U. Bowie, Backbone Hydrogen Bond Strengths Can Vary Widely in Transmembrane Helices. *J Am Chem Soc* **139**, 10742–10749 (2017).

81. A. K. Doura, F. J. Kobus, L. Dubrovsky, E. Hibbard, K. G. Fleming, Sequence context modulates the stability of a GxxxG mediated transmembrane helix-helix dimer. *J Mol Biol* **341**, 991–998 (2004).

82. M. T. Duong, T. M. Jaszewski, K. G. Fleming, K. R. MacKenzie, Changes in Apparent Free Energy of Helix-Helix Dimerization in a Biological Membrane Due to Point Mutations. *J Mol Biol* **371**, 422–434 (2007).

83. H. Hong, J. U. Bowie, Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments. *J. Am. Chem. Soc.* **133**, 11389–11398 (2011).

84. K. A. Gaffney, H. Hong, The rhomboid protease GlpG has weak interaction energies in its active site hydrogen bond network. *J. Gen. Physiol.* (2019) https:/doi.org/10.1085/jgp.201812047.

85. Y. C. Chang, J. U. Bowie, Measuring membrane protein stability under native conditions. *Proc Natl Acad Sci U S A* **111**, 219–224 (2014).

86. H. Hong, Y. C. Chang, J. U. Bowie, Measuring transmembrane helix interaction strengths in lipid bilayers using steric trapping. *Methods Mol. Biol.* **1063**, 37–56 (2013).

87. R. Guo, *et al.*, Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat Chem Biol* **12**, 353–360 (2016).

88. R. Chadda, J. L. Robertson, Measuring Membrane Protein Dimerization Equilibrium in Lipid Bilayers by Single-Molecule Fluorescence Microscopy. *Methods Enzym.* **581**, 53–82 (2016).

89. R. Chadda, *et al.*, The dimerization equilibrium of a ClC Cl(-)/H(+) antiporter in lipid bilayers. *Elife* **5** (2016).

90. R. Chadda, L. Cliff, M. Brimberry, J. L. Robertson, A model-free method for measuring

dimerization free energies of CLC-ec1 in lipid bilayers. *J. Gen. Physiol.* (2018) https:/doi.org/10.1085/jgp.201711893.

91. M. Mravic, *et al.*, Packing of apolar side chains enables accurate design of highly stable membrane proteins. *Science (80-.).* (2019) https:/doi.org/10.1126/science.aav7541.

92. Z. Cao, J. U. Bowie, An energetic scale for equilibrium H/D fractionation factors illuminates hydrogen bond free energies in proteins. *Protein Sci* **23**, 566–575 (2014).

93. M. J. Bower, F. E. Cohen, R. L. Dunbrack, Sidechain prediction from a backbone-dependent rotamer library: a new tool for homology modeling. *J. Mol. Biol.* **267**, 1268–1282 (1997).

94. E. S. O'Brien, *et al.*, Membrane Proteins Have Distinct Fast Internal Motion and Residual Conformational Entropy. *Angew. Chemie - Int. Ed.* (2020) https:/doi.org/10.1002/anie.202003527.

95. W. C. Wimley, T. P. Creamer, S. H. White, Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35**, 5109–5124 (1996).

96. A. Radzicka, R. Wolfenden, Comparing the Polarities of the Amino Acids: Side-Chain Distribution Coefficients between the Vapor Phase, Cyclohexane, 1-Octanol, and Neutral Aqueous Solution. *Biochemistry* **27**, 1664–1670 (1988).

97. C. P. Moon, K. G. Fleming, *Using tryptophan fluorescence to measure the stability of membrane proteins folded in liposomes*, 1st Ed. (Elsevier Inc., 2011).

98. C. P. Moon, S. Kwon, K. G. Fleming, Overcoming hysteresis to attain reversible equilibrium folding for outer membrane phospholipase A in phospholipid bilayers. *J. Mol. Biol.* **413**, 484–494 (2011).

99. W. C. Wimley, S. H. White, Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Mol. Biol.* **3**, 842–848 (1996).

100. M. B. Ulmschneider, M. S. P. Sansom, Amino acid distributions in integral membrane protein structures. *Biochim. Biophys. Acta - Biomembr.* **1512**, 1–14 (2001).

101. A. Jomaa, D. Boehringer, M. Leibundgut, N. Ban, Structures of the E. coli translating ribosome with SRP and its receptor and with the translocon. *Nat. Commun.* **7** (2016).

102. M. A. Lomize, A. L. Lomize, I. D. Pogozheva, H. I. Mosberg, OPM: Orientations of proteins in membranes database. *Bioinformatics* **22**, 623–625 (2006).

103. N. Noinaj, M. Guillier, T. J. Barnard,, S. K. Buchanan, TonB-Dependent Transporters: Regulation, Structure, and Function. *Annu. Rev. Microbiol.* **64**, 43–60 (2010).

104. R. E. Bishop, Structural biology of membrane-intrinsic β-barrel enzymes: Sentinels of the bacterial outer membrane. *Biochim. Biophys. Acta - Biomembr.* **1778**, 1881–1896 (2008).

105. R. E. Bishop, The lipid A palmitoyltransferase PagP: Molecular mechanisms and role in bacterial pathogenesis. *Mol. Microbiol.* **57**, 900–912 (2005).

106. T. J. Silhavy, J. C. Malinverni, Assembly of Outer Membrane β-Barrel Proteins: the Bam Complex. *EcoSal Plus* **4** (2011).

107. A. M. Plummer, K. G. Fleming, BamA Alone Accelerates Outer Membrane Protein Folding In Vitro through a Catalytic Mechanism. *Biochemistry* **54**, 6009–6011 (2015).

108. D. C. Marx, *et al.*, SurA is a cryptically grooved chaperone that expands unfolded outer membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* (2020) https:/doi.org/10.1073/pnas.2008175117 (October 27, 2020).

109. M. L. Johnson, J. J. Correia, D. A. Yphantis, H. R. Halvorson, Analysis of Data from the Analytical Ultracentrifuge by Nonlinear Least-Squares Techniques. *Biophys. J.* **36**, 575–588 (1981).

110. T. M. Laue, B. D. Shah, T. M. Ridgeway, S. L. Pelletier, "Computer-aided interpretation of

analytical sedimentation data for proteins" in *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, S. Harding, A. Rowe, J. Hoarton, Eds. (Royal Society of Chemistry, 1992), p. 90.

111. N. R. Zaccai, *et al.*, Deuterium Labeling Together with Contrast Variation Small-Angle Neutron Scattering Suggests How Skp Captures and Releases Unfolded Outer Membrane Proteins. *Methods Enzymol.* **566**, 159–210 (2016).

112. K. L. Sarachan, J. E. Curtis, S. Krueger, Small-angle scattering contrast calculator for protein and nucleic acid complexes in solution. *J. Appl. Crystallogr.* **46**, 1889–1893 (2013).

113. J. E. Curtis, S. Raghunandan, H. Nanda, S. Krueger, SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. *Comput. Phys. Commun.* **183**, 382–389 (2012).

114. A. Fiser, R. K. G. Do, A. Šali, Modeling of loops in protein structures. *Protein Sci.* (2000) https:/doi.org/10.1110/ps.9.9.1753.

115. SASSIE-Web, SASSIE-Web. *https://sassie-web.chem.utk.edu/sassie2/*.

116. J. Trewhella, *et al.*, 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: An update. *Acta Crystallogr. Sect. D Struct. Biol.* (2017) https:/doi.org/10.1107/S2059798317011597.

117. D. W. Bolen, M. M. Santoro, Unfolding Free Energy Changes Determined By The Linear Extrapolation Method. 2. Incorporation of Ag°n_u Values In A Thermodynamic Cycle. *Biochemistry* **27**, 8069–8074 (1988).

118. F. Stengel, *et al.*, Quaternary dynamics and plasticity underlie small heat shock protein chaperone function. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2007–12 (2010).

119. K. C. Giese, E. Vierling, Changes in oligomerization are essential for the chaperone activity

of a small heat shock protein in vivo and in vitro. *J. Biol. Chem.* **277**, 46310–46318 (2002).

120. L. Liu, J. Chen, B. Yang, Y. Wang, Oligomer-dependent and -independent chaperone activity of sHsps in different stressed conditions. *FEBS Open Bio* **5**, 155–162 (2015).

121. F. A. Saul, *et al.*, Structural and Functional Studies of FkpA from Escherichia coli, a cis/trans Peptidyl-prolyl Isomerase with Chaperone Activity. *J. Mol. Biol.* **335**, 595–608 (2004).

122. X. Ge, *et al.*, Identification of FkpA as a key quality control factor for the biogenesis of outer membrane proteins under heat shock conditions. *J. Bacteriol.* **196**, 672–680 (2014).

123. L. Arike, *et al.*, Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli. *J. Proteomics* **75**, 5437–5448 (2012).

124. T. Masuda, N. Saito, M. Tomita, Y. Ishihama, Unbiased quantitation of Escherichia coli membrane proteome using phase transfer surfactants. *Mol. Cell. Proteomics* **8**, 2770–2777 (2009).

125. P. J. Fleming, K. G. Fleming, HullRad: Fast Calculations of Folded and Disordered Protein and Nucleic Acid Hydrodynamic Properties. *Biophys. J.* **114**, 856–869 (2018).

126. M. M. Santoro, D. W. Bolen, Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **27**, 8063–8068 (1988).

127. C. W. Sandlin, N. R. Zaccai, K. G. Fleming, Skp Trimer Formation Is Insensitive to Salts in the Physiological Range. *Biochemistry* **54**, 7059–62 (2015).

128. B. M. Burmann, C. Wang, S. Hiller, Conformation and dynamics of the periplasmic membrane-protein-chaperone complexes OmpX-Skp and tOmpA-Skp. *Nat. Struct. Mol. Biol.* **20**, 1265–1272 (2013).

129. T. A. Walton, C. M. Sandoval, C. A. Fowler, A. Pardi, M. C. Sousa, The cavity-chaperone Skp protects its substrate from aggregation but allows independent folding of substrate domains. *Proc Natl Acad Sci U S A* **106**, 1772–1777 (2009).

130. K. Ramm, A. Pluckthun, The periplasmic Escherichia coli peptidylprolyl cis,trans-isomerase FkpA. II. Isomerase-independent chaperone activity in vitro. *J Biol Chem* **275**, 17106–17113 (2000).

131. S. Alvira, *et al.*, Trans-membrane association of the Sec and BAM complexes for bacterial outer-membrane biogenesis. *bioRxiv* https:/doi.org/10.1101/589077 (December 9, 2019).

132. J. A. Riback, *et al.*, Commonly used FRET fluorophores promote collapse of an otherwise disordered protein. *Proc. Natl. Acad. Sci. U. S. A.* (2019) https:/doi.org/10.1073/pnas.1813038116.

133. F. Chiti, C. M. Dobson, Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).

134. C. M. Dobson, Principles of protein folding, misfolding and aggregation in *Seminars in Cell and Developmental Biology*, (Elsevier Ltd, 2004), pp. 3–16.

135. J. M. Barral, S. A. Broadley, G. Schaffar, F. U. Hartl, Roles of molecular chaperones in protein misfolding diseases. *Semin. Cell Dev. Biol.* **15**, 17–29 (2004).

136. F. U. Hartl, M. Hayer-Hartl, Protein folding. Molecular chaperones in the cytosol: From nascent chain to folded protein. *Science (80-.).* **295**, 1852–1858 (2002).

137. Y. E. Kim, M. S. Hipp, A. Bracher, M. Hayer-Hartl, F. Ulrich Hartl, Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annu. Rev. Biochem.* **82**, 323–355 (2013).

138. H. Saibil, Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* **14**, 630–642 (2013).

139. F. U. Hartl, Molecular chaperones in cellular protein folding. *Nature* **381**, 571–580 (1996).

140. F. U. Hartl, A. Bracher, M. Hayer-Hartl, Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332 (2011).

141. T. Ulrich, D. Rapaport, Biogenesis of beta-barrel proteins in evolutionary context. *Int. J. Med. Microbiol.* **305**, 259–64 (2015).

142. J. M. Herrmann, J. Riemer, Three approaches to one problem: protein folding in the periplasm, the endoplasmic reticulum, and the intermembrane space. *Antioxid. Redox Signal.* **21**, 438–56 (2014).

143. A. Luther, *et al.*, Chimeric peptidomimetic antibiotics against Gram-negative bacteria. *Nature* **576**, 452–458 (2019).

144. E. M. Hart, *et al.*, A small-molecule inhibitor of BamA impervious to efflux and the outer membrane permeability barrier. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21748–21757 (2019).

145. Y. Imai, *et al.*, A new antibiotic selectively kills Gram-negative pathogens. *Nature* **576**, 459–464 (2019).

146. B. Schiffrin, *et al.*, Effects of Periplasmic Chaperones and Membrane Thickness on BamA-Catalyzed Outer-Membrane Protein Folding. *J. Mol. Biol.* **429**, 3776–3792 (2017).

147. B. Schiffrin, *et al.*, Skp is a multivalent chaperone of outer-membrane proteins. *Nat. Struct. Mol. Biol.* **23**, 786–793 (2016).

148. T. Krojer, *et al.*, Structural basis for the regulated protease and chaperone function of DegP. *Nature* **453**, 885–890 (2008).

149. E. Bitto, D. B. McKay, Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure* **10**, 1489–98 (2002).

150. D. C. Marx, M. J. Leblanc, A. M. Plummer, S. Krueger, K. G. Fleming, Domain interactions

determine the conformational ensemble of the periplasmic chaperone <scp>SurA</scp>. *Protein Sci.* **29**, 2043–2053 (2020).

151. E. W. Bell, E. J. Zheng, L. M. Ryno, Identification of inhibitors of the E. coli chaperone SurA using in silico and in vitro techniques. *Bioorg. Med. Chem. Lett.* **28**, 3540–3548 (2018).

152. Q. Chai, *et al.*, Diverse sequences are functional at the C-terminus of the E. coli periplasmic chaperone SurA. *Protein Eng. Des. Sel.* **27**, 111–116 (2014).

153. G. Krainer, *et al.*, Slow Interconversion in a Heterogeneous Unfolded-State Ensemble of Outer-Membrane Phospholipase A (2017) https:/doi.org/10.1016/j.bpj.2017.05.037 (March 6, 2020).

154. A. E. Tan, N. K. Burgess, D. S. DeAndrad, J. D. Marold, K. G. Fleming, Self-association of unfolded outer membrane proteins. *Macromol. Biosci.* **10**, 763–767 (2010).

155. E. J. Danoff, K. G. Fleming, Aqueous, Unfolded OmpA forms amyloid-like fibrils upon self-association. *PLoS One* **10**, e0132301 (2015).

156. S. C. Gill, P. H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326 (1989).

157. D. C. Marx, K. G. Fleming, Influence of Protein Scaffold on Side-Chain Transfer Free Energies. *Biophys. J.* **113**, 597–604 (2017).

158. J. J. Lee, D. S. Berns, Protein aggregation. The effect of deuterium oxide on large protein aggregates of C-phycocyanin. *Biochem. J.* **110**, 465–470 (1968).

159. A. V. Semenyuk, D. I. Svergun, GNOM - A program package for small-angle scattering data processing. *J. Appl. Crystallogr.* **24**, 537–540 (1991).

160. M. C. Chambers, *et al.*, A cross-platform toolkit for mass spectrometry and proteomics. *Nat.*

Biotechnol. (2012) https:/doi.org/10.1038/nbt.2377.

161. M. Götze, *et al.*, Automated assignment of MS/MS cleavable cross-links in protein 3d-structure analysis. *J. Am. Soc. Mass Spectrom.* **26**, 83–97 (2014).

162. B. Webb, A. Sali, Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinforma.* **2016**, 5.6.1-5.6.37 (2016).

163. , PyMOL. *PyMOL Mol. Graph. Syst. Version 1.3 Schrödinger, LLC.*

164. A. D. MacKerell, *et al.*, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

165. J. C. Phillips, *et al.*, Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

166. G. C. P. Van Zundert, *et al.*, The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* (2016) https:/doi.org/10.1016/j.jmb.2015.09.014.

167. L. Wang, A. Brock, B. Herberich, P. G. Schultz, Expanding the genetic code of Escherichia coli. *Science (80-.).* **292**, 498–500 (2001).

168. B. Hagan, J. V. Staros, *Azides and nitrenes: reactivity and utility* (Academic Press, 1984).

169. S. C. Reddington, *et al.*, Different photochemical events of a genetically encoded phenyl azide define and modulate GFP fluorescence. *Angew. Chemie - Int. Ed.* **52**, 5974–5977 (2013).

170. S. Willkomm, A. Zander, D. Grohmann, "Site-specific fluorescent labeling of argonaute for FRET-based bio-assays" in *Methods in Molecular Biology*, (Humana Press Inc., 2017), pp. 291–304.

171. J. W. Chin, *et al.*, Addition of p-azido-L-phenylalanine to the genetic code of Escherichia

coli. *J. Am. Chem. Soc.* **124**, 9026–9027 (2002).

172. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

173. E. J. Danoff, K. G. Fleming, The soluble, periplasmic domain of OmpA folds as an independent unit and displays chaperone activity by reducing the self-association propensity of the unfolded OmpA transmembrane β-barrel. *Biophys. Chem.* **159**, 194–204 (2011).

174. S. El-Gebali, *et al.*, The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).

175. J. M. A. Bullock, J. Schwab, K. Thalassinos, M. Topf, The importance of non-accessible crosslinks and solvent accessible surface distance in modeling proteins with restraints from crosslinking mass spectrometry. *Mol. Cell. Proteomics* **15**, 2491–2500 (2016).

176. J. M. A. Bullock, K. Thalassinos, M. Topf, Jwalk and MNXL web server: model validation using restraints from crosslinking mass spectrometry. *Bioinformatics* **34**, 3584–3585 (2018).

177. F. Pedregosa, *et al.*, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

178. T. Saio, X. Guan, P. Rossi, A. Economou, C. G. Kalodimos, Structural Basis for Protein Antiaggregation Activity of the Trigger Factor Chaperone https:/doi.org/10.1126/science.1250494 (December 4, 2019).

179. A. J. Wirth, M. Gruebele, Quinary protein structure and the consequences of crowding in living cells: Leaving the test-tube behind. *BioEssays* **35**, 984–993 (2013).

180. S. Wu, *et al.*, Interaction between bacterial outer membrane proteins and periplasmic quality control factors: a kinetic partitioning mechanism. *Biochem. J.* **438**, 505–11 (2011).

181. D. Gessmann, *et al.*, Outer membrane β-barrel protein folding is physically controlled by periplasmic lipid head groups and BamA. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5878–5883 (2014).

182. M. Struyvé, M. Moons, J. Tommassen, Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. *J. Mol. Biol.* **218**, 141–8 (1991).

183. H. De Cock, M. Struyvé, M. Kleerebezem, T. Van Der Krift, J. Tommassen, Role of the carboxy-terminal phenylalanine in the biogenesis of outer membrane protein PhoE of Escherichia coli K-12. *J. Mol. Biol.* **269**, 473–478 (1997).

184. L. Maguire, M. Stefferson, M. D. Betterton, L. E. Hough, Design principles of selective transport through biopolymer barriers. *Phys. Rev. E* **100** (2019).

185. H. Gong, P. J. Fleming, G. D. Rose, Building native protein conformation from highly approximate backbone torsion angles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 16227–16232 (2005).

186. T. Kortemme, A. V. Morozov, D. Baker, An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.* **326**, 1239–1259 (2003).

187. G. Fiorin, M. L. Klein, J. Hénin, Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **111**, 3345–3362 (2013).

188. M. J. Graham, C. Combe, L. Kolbowski, J. Rappsilber, xiView: A common platform for the downstream analysis of Crosslinking Mass Spectrometry data. *bioRxiv*, 1–5 (2019).

189. H. Hong, Toward understanding driving forces in membrane protein folding. *Arch. Biochem. Biophys.* **564**, 297–313 (2014).

190. C. Tanford, The Hydrophobic Effect and the Organization of Living Matter. *Science* **200**,

1012–1018 (1978).

191. S. H. White, W. C. Wimley, Membrane protein folding and stability: Physical principles. *Ann. Rev. Biophys. Biomol. Str.* **28**, 319–365 (1999).

192. B. M. Kroncke, *et al.*, Documentation of an Imperative to Improve Methods for Predicting Membrane Protein Stability. *Biochemistry* **55**, 5002–5009 (2016).

193. K. Cox, M. S. P. Sansom, One membrane protein, two structures and six environments: A comparative molecular dynamics simulation study of the bacterial outer membrane protein PagP. *Mol. Membr. Biol.* **26**, 205–214 (2009).

194. J. A. Cuesta-Seijo, *et al.*, PagP crystallized from SDS/Cosolvent reveals the route for phospholipid access to the hydrocarbon ruler. *Structure* **18**, 1210–1219 (2010).

195. P. M. Hwang, *et al.*, Solution structure and dynamics of the outer membrane enzyme PagP by NMR. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 13560–13565 (2002).

196. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).

197. S. Jo, T. Kim, W. Im, Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS One* **2** (2007).

198. B. R. Brooks, *et al.*, CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).

199. E. L. Wu, *et al.*, CHARMM-GUI *Membrane Builder* toward realistic biological membrane simulations. *J. Comput. Chem.* **35**, 1997–2004 (2014).

200. J. Lee, *et al.*, CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **12**, 405–413 (2016).

201. M. A. Khan, R. E. Bishop, Molecular mechanism for lateral lipid diffusion between the outer membrane external leaflet and a β-barrel hydrocarbon ruler. *Biochemistry* **48**, 9745–9756 (2009).

202. M. M. Santoro, D. W. Bolen, A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range. *Biochemistry* **31**, 4901–4907 (1992).

203. C. Chothia, Hydrophobic bonding and accessible surface area in proteins. *Nature* **248**, 338–339 (1974).

204. D. Eisenberg, A. D. Mclachlan, Solvation energy in protein folding and binding. *Nature* **319**, 199–203 (1986).

205. A. Elazar, *et al.*, Mutational scanning reveals the determinants of protein insertion and association energetics in the plasma membrane. *Elife* **5** (2016).

206. M. W. MacArthur, J. M. Thornton, Influence of proline residues on protein conformation. *J. Mol. Biol.* **218**, 397–412 (1991).

207. S. Yohannan, *et al.*, Proline substitutions are not easily accommodated in a membrane protein. *J. Mol. Biol.* **341**, 1–6 (2004).

208. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33- (1996).

209. P. M. Hwang, R. E. Bishop, L. E. Kay, The integral membrane enzyme PagP alternates between two dynamically distinct states. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9618–9623 (2004).

210. J. D. Bloom, S. T. Labthavikul, C. R. Otey, F. H. Arnold, Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5869–5874 (2006).

211. M. Osadchy, R. Kolodny, Maps of protein structure space reveal a fundamental relationship between protein structure and function. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 12301–12306 (2011).

212. R. Reis, I. Moraes, Structural biology and structure–function relationships of membrane proteins. *Biochem. Soc. Trans.* **47**, 47–61 (2018).

213. E. Freire, The thermodynamic linkage between protein structure, stability, and function. *Methods Mol. Biol.* **168**, 37–68 (2001).

214. A. Goldenzweig, S. J. Fleishman, Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **87**, 105–129 (2018).

215. M. S. Almén, K. J. V. Nordström, R. Fredriksson, H. B. Schiöth, Mapping the human membrane proteome: A majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol.* **7**, 50 (2009).

216. D. L. Parton, J. W. Klingelhoefer, M. S. P. Sansom, Aggregation of model membrane proteins, modulated by hydrophobic mismatch, membrane curvature, and protein class. *Biophys. J.* **101**, 691–699 (2011).

217. T. Hessa, *et al.*, Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026–1030 (2007).

218. J. P. Ulmschneider, J. C. Smith, S. H. White, M. B. Ulmschneider, The importance of the membrane interface as the reference state for membrane protein stability. *Biochim. Biophys. Acta - Biomembr.* **1860**, 2539–2548 (2018).

219. S. Damodaran, K. B. Song, The role of solvent polarity in the free energy of transfer of amino acid side chains from water to organic solvents. *J. Biol. Chem.* **261**, 7220–7222 (1986).

220. M. C. Wiener, S. H. White, Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. II. Distribution and packing of terminal methyl groups. *Biophys. J.* **61**, 428–433 (1992).

221. M. C. Wiener, S. H. White, Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. III. Complete structure. *Biophys. J.* **61**, 434–447 (1992).

222. L. Wesson, D. Eisenberg, Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* **1**, 227–35 (1992).

223. K. Öjemalm, *et al.*, Apolar surface area determines the efficiency of translocon-mediated membrane-protein integration into the endoplasmic reticulum. *Proc Natl Acad Sci U S A* **108**, E359-364 (2011).

224. J. A. Reynolds, D. B. Gilbert, C. Tanford, Empirical Correlation Between Hydrophobic Free Energy and Aqueous Cavity Surface Area. *Proc. Natl. Acad. Sci.* **71**, 2925–2927 (1974).

225. T. Giorgino, Computing 1-D atomic densities in macromolecular simulations: The density profile tool for VMD. *Comput. Phys. Commun.* **185**, 317–322 (2014).

226. M. Lin, D. Gessmann, H. Naveed, J. Liang, Outer Membrane Protein Folding and Topology from a Computational Transfer Free Energy Scale. *J. Am. Chem. Soc.* **138**, 2592–2601 (2016).

227. M. J. M. Niesen, A. Müller-Lucks, R. Hedman, G. von Heijne, T. F. Miller, Forces on Nascent Polypeptides during Membrane Insertion and Translocation via the Sec Translocon. *Biophys. J.* **115**, 1885–1894 (2018).

228. S. Capponi, M. Heyden, A.-N. Bondar, D. J. Tobias, S. H. White, Anomalous behavior of water inside the SecY translocon. *Proc. Natl. Acad. Sci.*, 201424483 (2015).

229. R. F. Alford, P. J. Fleming, K. G. Fleming, J. J. Gray, Protein Structure Prediction and Design in a Biologically Realistic Implicit Membrane. *Biophys. J.* **118**, 2042–2055 (2020).

230. R. F. Alford, *et al.*, An Integrated Framework Advancing Membrane Protein Modeling and Design. *PLoS Comput. Biol.* **11** (2015).

231. J. Y. Weinstein, A. Elazar, S. J. Fleishman, A lipophilicity-based energy function for membrane-protein modelling and design. *PLoS Comput. Biol.* **15** (2019).

232. J. P. Overington, B. Al-Lazikani, A. L. Hopkins, How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).

233. D. Avci, M. K. Lemberg, Clipping or Extracting: Two Ways to Membrane Protein Degradation. *Trends Cell Biol.* **25**, 611–622 (2015).

234. S. I. Sukharev, W. J. Sigurdson, C. Kung, F. Sachs, Energetic and spatial parameters for gating of the bacterial large conductance mechanosensitive channel, MscL. *J. Gen. Physiol.* **113**, 525–539 (1999).

235. E. S. Haswell, R. Phillips, D. C. Rees, Mechanosensitive channels: What can they do and how do they do it? *Structure* **19**, 1356–1369 (2011).

236. S. Hofmann, *et al.*, Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature* **571**, 580–583 (2019).

237. J. T. Marinko, *et al.*, Folding and Misfolding of Human Membrane Proteins in Health and Disease: From Single Molecules to Cellular Proteostasis. *Chem. Rev.* **119**, 5537–5606 (2019).

238. W.-M. Yau, W. C. Wimley, K. Gawrisch, S. H. White, The Preference of Tryptophan for Membrane Interfaces †. *Biochemistry* **37**, 14713–14718 (1998).

239. D. C. Marx, K. G. Fleming, Local Hydrophobicity Modulates Membrane Protein Stability.

BioRxiv.org (2020) https:/doi.org/https://doi.org/10.1101/2020.09.01.277897.

Curriculum Vitae

Dagan Marx

T.C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, MD 21218

email: dmarx2@jhu.edu Phone: 352-682-1206

Born: January 23, 1992 in Iowa City, Iowa

EDUCATION:

08/14- 12/20 (Anticipated*) Ph.D. Candidate, Molecular Biophysics

Krieger School of Arts and Sciences, Johns Hopkins University

08/10- 05/14 B.S., Biochemistry and Environmental Science

College of Liberal Arts and Sciences, University of Iowa

PUBLICATIONS:

- **Marx, D.C.** and Fleming K.G. (2020). Local Bilayer Hydrophobicity Modulates Membrane Protein Stability. (*Submitted to JACS)* bioRxiv: https://doi.org/10.1101/2020.09.01.277897
- Vorobieva, A.A., White, P., Liang, B., Horne, J.E., Bera, A.K., Chow, C.M., Gerben, S., Marx, S., Kang, A., Stiving, A.Q., Harvey, S.R., **Marx, D.C.,** Khan, D.G., Fleming, K.G., Wysocki, V.H., Brockwell, D.J., Tamm, L.K., Radford S.E., Baker, D. (2020) *De novo* Design of Transmembrane β-barrels. (*Submitted to Nature*) bioRxiv: https://doi.org/10.1101/2020.10.22.346965
- **Marx, D.C.**, Plummer, A.M., Faustino, A.M., Roskopf, M.A., Leblanc, M.J., Devlin, T., Lessen, H.J., Amann, B.T., Fleming, P.J., Krueger, S., Fried, S.D., and Fleming, K.G. (2020). SurA is a Cryptically Grooved Chaperone That Expands Unfolded Outer Membrane Proteins. *PNAS.* 117(45):28026-28035. doi: 10.1073/pnas.2008175117.

- **Marx, D.C.,** Leblanc, Plummer, A.M., Krueger, S., Fleming, K.G. (2020). Domain Interactions Determine the Conformational Ensemble of the Periplasmic Chaperone SurA. *Protein Science*. 29(10)2043-2053 doi:10.1002/pro.3924
- **Marx, D.C.** and Fleming, K.G. (2017). Influence of Protein Scaffold on Side Chain Transfer Free Energies. *Biophysical Journal*. 13(3):597-604. doi: 10.1016/j.bpj.2017.06.032.
- Hovey, L., Fowler, C.A., Mahling, R., Lin, Z., Miller, M.S., **Marx, D.C.**, Yoder, J.B., Kim, E.H., Tefft, K.M., Waite, B.C., Feldkamp, M.D., Yu, L., Shea, M.A. (2017). Calcium triggers reversal of calmodulin on nested anti-parallel sites in the IQ motif of the neuronal voltage-dependent sodium channel NaV1.2. *Biophysical Chemistry.* 224:1-19. doi: 10.1016/j.bpc.2017.02.006.

RESEARCH EXPERIENCE:

Graduate Research Assistant, Lab of Dr. Karen G. Fleming (05/15-present)

Department of Biophysics, Johns Hopkins University, Baltimore, MD

Project: Integrative Characterization of SurA Interaction with an Unfolded Membrane Protein Client

- Structurally characterized the complex formed between the E. *coli* periplasmic chaperone SurA and unfolded, client outer membrane proteins. Combined chemical crosslinking, small angle neutron scattering, mass spectrometry, and molecular modeling to build the first structural models of this unorthodox chaperone-membrane protein complex

Project: Influence of Protein Scaffold and Bilayer Position on Side Chain Transfer Free Energies

- Measured side chain transfer free energies from water-to-bilayer using a previously developed fluorescence-based host-guest methodology, combined with molecular dynamics simulations. Determined a hydrophobicity scale for all twenty amino acids in the protein scaffold PagP. Determined the bilayer position dependence of non-polar side chain transfer free energies using OmpLA as a scaffold, finding a linear correlation between nonpolar solvation parameter and water concentration in the bilayer.

Project: Client Binding Energetics and Kinetics to the Periplasmic Chaperone FkpA

- Measured and modeled the dimerization constant of FkpA, as well as the binding constant for FkpA to an unfolded outer membrane protein client using AUC and molecular dynamics simulations. Using a gel-shift based folding assay, found novel chaperone function and kinetics of FkpA compared to other periplasmic chaperones.

Mentor to Undergraduate Researchers: Michaela Roskopf, Mathis Leblanc, and Mark Kreutzberger

- Michaela's project was to work in concert with me on the experimental and computation aspects of the characterization of FkpA's chaperone functions. Mathis' project was to experimentally investigate the effects of conformational dynamics on SurA function. Mark worked to express, purify, and fold variants of OmpLA used to measure transfer free energies.

- As a mentor I designed experimental and computational aspects of each project, conducted bi-weekly meetings to further understanding about project direction and relevant literature, and trained students in experimental techniques and communication skills

Undergraduate Research Assistant, Lab of Dr. Madeline A. Shea (01/11-07/14)

Department of Biochemistry, University of Iowa, Iowa City, IA

Project: Calmodulin Interaction with hNa$_V$ IQ motifs

- Measured binding affinities of calmodulin (+/- Ca^{2+}) to a C-terminal calmodulin binding motif in the nine human isoforms of voltage-gated sodium channels using a FRET biosensor. Found two classes of channels that differ based on their binding affinity with calmodulin, which is implicated in channel inactivation.

PRESENTATIONS:

Invited Talks (n = 5):

- **Marx, D.C.** (2021, June/July). Gordon Research Conference on Membrane Protein Folding. Castelldefels, Barcelona, Spain.
- **Marx, D.C.** (2020, February). *Towards Understand the Effect of Water on Membrane Protein Stability*. Biophysical Society 64th Annual Meeting, San Diego, CA. Platform: Membrane Protein Dynamics and Folding I.
- **Marx, D.C.** (2019, August). *Membrane Position Dependence of the Nonpolar Solvation Parameter.* Institute for Biophysical Research. Baltimore, MD.
- **Marx, D.C.** (2017, October). Student Invited Speaker, JHU PMB Student Evening Series.

- **Marx. D.C.** (2016, September). *Side Chain Hydrophobicity Scale Using the Tilted Beta-Barrel PagP.* Saturday Night Thermo Session @ 30th Annual Gibbs Conference on Biothermodynamics, Carbondale, IL.
- **Marx, D.C.** (2014, September). *Breaking the IQ code: deciphering the keys for tight CaM-NaV interaction.* 28th Annual Gibbs Conference on Biothermodynamics, Carbondale, IL.

Posters (n = 18):

- **Marx, D.C.**, Plummer, A.M., Faustino, A.M., Roskopf, M.A., Leblanc, M.J., Devlin, T., Lessen, H.J., Majumdar, A., Amann, B.T., Fleming, P.J., Krueger, S., Fried, S.D., and Fleming, K.G. (2020, February). *SurA is a "Groove-y" Chaperone That Expands Unfolded Outer Membrane Proteins.* Poster presented at the Biophysical Society 64th Annual Meeting, San Diego, CA.
- Roskopf, M.A, **Marx, D.C.,** Fleming, K.G. (2020, February). *Energetics of Dimeric FkpA Binding to a Native Unfolded Membrane Protein Client.* Poster presented at the Biophysical Society 64th Annual Meeting, San Diego, CA.
- **Marx, D.C.,** and Fleming K.G. (2019, October) *Bilayer Depth Dependence of Hydrophobic Amino Acid Transfer Free Energies.* Poster presented at the 33rd Annual Gibbs Conference on Biothermodynamics, Carbondale, IL.
- Roskopf, M.A., **Marx, D.C.,** Plummer, A.M., Fleming, K.G. (2019, October). *Energetics of Dimeric FkpA Binding to a Native Unfolded Membrane Protein Client.* Poster presented at the 33rd Annual Gibbs Conference on

Biothermodynamics, Carbondale, IL.

- **Marx, D.C.,** and Fleming K.G. (2019, July) *Bilayer Depth Dependence of Hydrophobic Amino Acid Transfer Free Energies.* Gordon Research Conference: Membrane Protein Folding, Easton, MA.
- **Marx, D.C.,** and Fleming K.G. (2019, March) *Bilayer Depth Dependence of Hydrophobic Amino Acid Transfer Free Energies.* Poster presented at the Biophysical Society 63rd Annual Meeting, Baltimore, MD.
- Roskopf, M.A, **Marx, D.C.,** Plummer, A.M., Bubb, Q.R., Fleming, K.G. (2019, March). *Energetics of Dimeric FkpA Binding to a Native Unfolded Membrane Protein Client.* Poster presented at the Biophysical Society 63re Annual Meeting, Baltimore, MD.
- **Marx, D.C.,** and Fleming K.G. (2018, October) *Bilayer Depth Dependence of Hydrophobic Amino Acid Transfer Free Energies.* Poster presented at the 32nd Annual Gibbs Conference on Biothermodynamics, Carbondale, IL.
- **Marx, D.C.,** and Fleming K.G. (2017, June) *Side Chain Hydrophobicity Scale Using the*

Tilted Beta-Barrel PagP. Poster presented at the Gordon Research Conference: Membrane Protein Folding, Easton, MA.

- **Marx, D.C.,** and Fleming K.G. (2017, February) *Side Chain Hydrophobicity Scale Using the Tilted Beta-Barrel PagP.* Poster presented at the Biophysical Society 61st Annual Meeting, New Orleans, LA.

- **Marx. D.C.** and Fleming K.G. (2016, September). *Side Chain Hydrophobicity Scale Using the Tilted Beta-Barrel PagP.* Poster presented at the 30th Annual Gibbs Conference on Biothermodynamics, Carbondale, IL.
- Hovey L., Andresen, C., **Marx, D.C.,** Shea, M.A. (2015, February) *Recruitment of Calmodulin to the Tail of the Voltage-Gated Sodium Channel Nav1.2.* Poster presented at the Biophysical Society 59th Annual Meeting, Baltimore, MD.
- **Marx, D.C.,** Miller, M.S., Hovey, L., Tefft, K.M., Yoder, J.B., Kim, E., Martin S.C., Feldkamp, M.D., Shea, M.A. (2014, February). *Specificity of Calmodulin Recognition of Human Voltage Gated Sodium Channels.* Poster presented at the Biophysical Society 58th Annual Meeting, San Francisco, CA.
- Hovey, L., Miller, M.S., **Marx, D.C.,** Tefft, K.M., Kim, E., Yoder, J.B., Shea, M.A. (2014, February). *Determinants of Preferential Binding of Apo Calmodulin to the IQ Motif of the Neuronal Sodium Channel Nav1.2.* Poster presented at the Biophysical Society 58th Annual Meeting, San Francisco, CA.
- **Marx, D.C.,** Tefft, K., Miller, M.S., Yoder, J.B., Hovey, L., Kim, E., Martin, S.C., Tarleton, A., Waite, B.C., Feldkamp, M.D., Shea, M.A. (2013, October). *Bifurcation of Human Sodium Channels based on Calmodulin Discrimination of IQ Motifs.* Poster presented at the 27th Annual Gibbs Conference on Biothermodynamics, Carbondale, IL.
- Shea M.A., Miller, M.S., Yoder, J.B., Martin, S.C., **Marx, D.C.,** Kim, E., Tarleton, D., Miller, E.S., Waite, B.C., Wold, A.O., Feldkamp, M.D. (2013, February). *Calmodulin Discrimination Between Voltage-Dependent Sodium Channel IQ*

Motifs. Poster presented at the Biophysical Society 57th Annual Meeting, Philadelphia, PA.

- **Marx, D.C.,** Kim, E., Miller, M.S., Yoder, J.B., Martin, S.C., Tarleton, A., Shea, M.A. (2012, September). *Calmodulin Recognition of Voltage-Gated Sodium Channels in Excitable Cells.* Poster presented at the 26th Annual Gibbs Conference on Biothermodynamics, Carbondale, IL.
- Miller, M.S., Yoder, J.B., Martin, S.C., Waite, B.C., **Marx, D.C**., Kim, E., Miller, E.S., Tarleton, D., Feldkamp, M.D., Shea, M.A. (2011, September) *Recognition of Voltage-Dependent Sodium Channels by Calmodulin.* Poster presented at the 25th Annual Gibbs Conference on Biothermodynamics, Carbondale, IL.

TEACHING EXPERIENCE:

02/20 Co-instructor: Advanced Seminar in Membrane Protein Structure, Function, and Pharmacology. Johns Hopkins University, Baltimore, MD

01/18-05/19 Teaching Assistant: Protein Engineering and Biochemistry Lab (Undergraduate Course). Johns Hopkins University, Baltimore, MD

08/15-12/15 Teaching Assistant: Proteins and Nucleic Acids (Graduate Course) Johns Hopkins University, Baltimore, MD